DOCUMENT RESUME

ED 096 339 TM 003 945

AUTHOR	McBride, James R.; Weiss, David J.
TITLE	A Word Knowledge Item Pool for Adaptive Ability Measurement. Research Report No. 74-2.
INSTITUTION	Minnesota Univ., Minneapolis. Dept. of Psychology.
SPONS AGENCY	Office of Naval Research, Washington, D.C. Personnel and Training Research Programs Office.
PUB DATE	Jun 74
NOTE	92p.
EDRS PRICE	MF-$0.75 HC-$4.20 PLUS POSTAGE
DESCRIPTORS	College Students; Internal Scaling; *Item Analysis; *Item Banks; Norms; *Test Construction; *Vocabulary Skills
IDENTIFIERS	*Adaptive Ability Measurement; Computer Assisted Testing

ABSTRACT

A series of four vocabulary norming tests was used to develop a large, homogeneous pool of vocabulary test items for use in computer-administered adaptive testing research. Five hundred seventy-five unique vocabulary knowledge items were divided among the four norming tests, and administered to separate groups of college undergraduates. Norming tests were administered by computer or paper and pencil in fixed and random order. Analyses showed no effects due to item order or mode of administration. Item difficulty and discrimination indices of both the classical test model and the normal ogive item model were derived on the norming data. On the basis of item analysis results, 369 items were selected as satisfactory for the adaptive testing item pool. Factor-analytic studies of subsets of the 369 items confirmed the assumption of unidimensionality of the selected item pool. On the basis of known technical limitations in the research and the unique problems of developing item pools for adaptive testing, an outline was developed for the design of future norming studies specifically intended to develop large homogeneous pools of test items for use in computer-administered adaptive ability measurement. (Author)

ED 096339

A WORD KNOWLEDGE ITEM POOL FOR ADAPTIVE ABILITY MEASUREMENT

James R. McBride

and

David J. Weiss

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Research Report 74-2

Psychometric Methods Program
Department of Psychology
University of Minnesota
Minneapolis, MN 55455

June 1974

TM 003 945

Prepared under contract No. N00014-67-A-0113-0029
NR No. 150-343, with the Personnel and
Training Research Programs, Psychological Sciences Division
Office of Naval Research

Approved for public release; distribution unlimited.
Reproduction in whole or in part is permitted for
any purpose of the United States Government.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report 74-2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Word Knowledge Item Pool for Adaptive Ability Measurement		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) James R. McBride and David J. Weiss		8. CONTRACT OR GRANT NUMBER(s) N00014-67-0113-0029
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, Minnesota 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.: 61153N PROJ: RR042-04 T.A.: RR042-04-01 W.U.: NR150-343
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, VA 22217		12. REPORT DATE June 1974 13. NUMBER OF PAGES 81
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

testing	sequential testing	programmed testing
ability testing	branched testing	response-contingent testing
computerized testing	individualized testing	automated testing
adaptive testing	tailored testing	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A series of four vocabulary norming tests was used to develop a large, homogeneous pool of vocabulary test items for use in computer-administered adaptive testing research. 575 unique vocabulary knowledge items were divided among the four norming tests, and administered to separate groups of college undergraduates. Norming tests were administered by computer or paper and pencil in fixed and random orders. (over)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20 (cont'd)

Analyses showed no effects due to item order or mode of administration. Item difficulty and discrimination indices of both the classical test model and the normal ogive item model were derived on the norming data. On the basis of item analysis results, 369 items were selected as satisfactory for the adaptive testing item pool. Factor-analytic studies of subsets of the 369 items confirmed the assumption of unidimensionality of the selected item pool. On the basis of known technical limitations in the research and the unique problems of developing item pools for adaptive testing, an outline was developed for the design of future norming studies specifically intended to develop large homogeneous pools of test items for use in computer-administered adaptive ability measurement.

Contents

Introduction 1

Mode of administration and item order effects 3
Rationale 3
Method 5
Subjects 5
Instruments 7
Item format 7
Conditions of administration 7
Item sources and item construction ... 8
Experimental variables 9
Data collection 9
Data analysis 11
Test medium and item order effects ... 11
Group differences in ability 13
Results 13
Test medium and item order effects 13
Group ability effects 13
Analysis of calibration tests 16
Ability level comparison 16
Discussion 19

Development of the final item pool 20
Rationale 20
Method 22
Item parameter estimation 22
Traditional item parameters 22
Normal ogive item parameters 24
Selection of the final item pool 25
Results 26
Traditional item parameters 26
Normal ogive item parameters 26
Description of the final item pool..... 29

Dimensionality of the final item pool 29
Rationale 29
Method 32
Results 34
Factor contributions 34
Factor structure 37
Item discriminations and first factor loadings 37
Summary 37

Limitations of the final item pool..... 40
Generalizability 40
Traditional item parameters 41
Normal ogive item parameters 42
Factor analyses 42

A design for future norming 43
Basic steps 44
Operationalization 44
Population definition 44
Initial norming 45
Development of the calibration test 45
Development of a trial item pool..... 47
Norming of the trial item pool..... 47
Item analyses 48
Selection of the final item pool..... 48

References 49

Appendix A: Traditional and normal ogive parameters for the items of the norming item pool 51

Appendix B: Cross-tabulation of the 369 items selected for the final item pool by traditional difficulty and discrimination indices 62

Appendix C: Matrices of tetrachoric item intercorrelations and principal factor loadings tables for the six real data and six random data subtests 64

A WORD KNOWLEDGE ITEM POOL FOR ADAPTIVE ABILITY MEASUREMENT

Administration of ability and achievement tests by computer, rather than by an individual in a clinical setting or by paper and pencil in a group-test setting, offers promise of eliminating some disadvantages of each of the latter methods, while incorporating the advantages of each. Specifically, a test administered by a computer terminal--teletype or cathode-ray-terminal (CRT)--has promise of alleviating effects due to examiner variables (race, sex, rapport), answer sheet arrangement, and subjective item scoring. At the same time, such a test offers the advantages of standardized item presentation, recording of response latency, relatively inexpensive administration and immediate scoring (Weiss & Betz, 1973).

Furthermore, computer administration makes possible the use of adaptive testing strategies, previously limited to individual tests such as the Stanford-Binet, offering flexible administration without the possibly distorting effects on test scores due to a human test administrator. Adaptive testing, in which the difficulty of the test is adjusted to the testee's ability level while testing is in progress, has begun to demonstrate higher levels of both reliability and validity in comparison to conventional tests (Weiss & Betz, 1973). At the same time, adaptive tests offer the additional advantage of eliminating speed effects in test scores. On the other hand, adaptive tests have the disadvantage of requiring a relatively large pool of test items, in comparison with the number of items in a standardized group or individual test.

The purpose of the present paper is to report on the development of an item pool for use in the investigation of the relative merits of several different strategies of computer-administered adaptive ability tests. For the purpose of this research it was desired to have a pool of items having a relatively univocal factor structure, yet measuring an empirically significant dimension of ability. Since tests of vocabulary knowledge have repeatedly been shown to satisfy both of those criteria (Lord & Novick, 1968, p. 381), the pool was developed using multiple-choice vocabulary items. The final item pool was designed to meet the following criteria: 1) it would be sufficiently large to allow several independent (non-overlapping) subtests to be drawn from its items; 2) the items would span the entire range of item difficulty relative to the population of interest; and 3) it would consist of highly discriminating items.

The research reported in this paper was concerned with assembling such a large item pool, norming its items relative to one population of interest, estimating item parameters of difficulty and discrimination within that population, and investigating the hypothesis that the items selected as

satisfactory for inclusion in the final item pool essentially measured a single dimension of ability. The parent research project will evaluate strategies of computer-administered testing on several independent groups of persons, but the first phase of this evaluation (e.g., Betz & Weiss, 1973; Larkin & Weiss, 1974) utilized a readily accessible population--students enrolled in several psychology courses at the University of Minnesota.

The first stage of item-pool development involved norming a large number of vocabulary items on sufficient numbers of students to yield stable estimates of both their traditional difficulty and discrimination indices, and of the parameters of their normal ogive item characteristic curves. Later stages of this development normed items on groups other than psychology students, including high school students, junior college students, and vocational rehabilitation clients; details of these studies will be reported separately.

Since the adaptive tests in which the selected items were to appear are administered at a computer terminal (DeWitt & Weiss, 1974), it is appropriate that norming tests also be computer-administered. In the interest of economy, however, much of the norming was accomplished by means of paper-and-pencil tests. Thus, another research objective was to determine if there were differences in item parameters between CRT and paper-and-pencil administration, and if necessary, to account for those differences when estimating item parameters. Yet another variable to be accounted for was item arrangement, or order of presentation, since several methods of item arrangement were used at different points in the norming study.

The first section of this report presents the results of four norming test administrations under varying sets of conditions. The analysis is concerned with the problem of determining whether norming tests administered in different modes (pencil-and-paper vs. CRT), and containing items in different orders, resulted in different estimates of item difficulties and discriminations. These analyses were designed to determine the most feasible method of norming test items for use in computer-administered tests.

The second part of this report describes the method by which two sets of item parameter estimates were obtained for every item in the pool. The first set of parameters consisted of the traditional difficulty and discrimination item parameters. The second set of item parameters estimated were the parameters of the normal ogive item characteristic curve for each item. Interest in these parameters reflects the need for item parameters which are invariant across groups differing in average ability. Such properties of invariance are required

for subsequent research utilizing the same item pool with subject groups differing in ability from the norming group. Another reason for the desirability of the normal ogive parameters is their usefulness in item selection. Since in item characteristic curve theory equal difficulty-scale intervals represent equal intervals of the underlying ability, knowledge of the normal ogive parameters makes it theoretically possible to select test items to yield any desired distribution of test scores if certain characteristics of the testee group are known (Lord & Novick, 1968, pp. 386-392).

Also described in the second part of the report are the criteria on the basis of which items from the initial development item pool were selected for inclusion in the final item pool, and the composition of the final item pool.

The third part of this report is concerned with the investigation of the dimensionality of the set of test items selected to comprise the final item pool. The issue of dimensionality of the item pool is an important one, since it determines the appropriateness of the use of item characteristic curve parameters for item scaling, as well as the appropriateness of certain adaptive testing models using a given pool of test items.

The report concludes with a discussion of some technical limitations of the present study which have implications for the generality of the results reported. Included is a discussion of considerations peculiar to developing an item pool for computer-administered testing, and an outline of a research design for a comprehensive and large-scale norming study for the future development of item pools to be used in adaptive ability measurement.

MODE OF ADMINISTRATION AND ITEM ORDER EFFECTS

Rationale

In the course of data collection for the present research, 575 unique vocabulary test items have been incorporated in one or more of four tests. Each of the four tests differed from the others in terms of its mode of administration, or the ordering of items within the test, or both. Additionally, each test was administered at a different time, and to a different group of examinees. Each point of difference from one test to another represents a potential source of extraneous variance in the difficulty and discrimination estimates determined for each item.

The purpose of the analyses described here was to investigate whether any differential effects on overall test performance occurred due to differences in mode of administration, order of item presentation, or ability of the examinee group. If no such differential effects were found, then for each test the observed item difficulty values can be taken as estimates of the difficulty of each item within the experimental population. If, however, it was found that test performance covaried with differences in the three variables in question, then statistical adjustments to the observed item statistics would be necessary so that the item parameter estimates, obtained on the basis of different test administrations, would be directly comparable.

Should it be found, for instance, that test results obtained by means of paper-and-pencil administration differed markedly from those obtained from computer administration, then the paper-and-pencil results would have to be corrected in some manner for use in computer-administered tests. If comparability was not observed, and if corrections were not feasible, then all norming efforts would have to be done using the computer mode of administration so that they could be used in computer-administered adaptive tests.

Another administration variable whose effects needed to be identified was that of item order within the test. In most standardized tests item order does not vary; under computer administration items are presented to different testees in different orders, both within and between strategies of adaptive testing (Weiss, 1974). Whether differences between fixed and varying order of item presentation affected overall examinee performance (and hence item analysis results) needed to be determined because of its import for adaptive testing, and in the present instance because of its import for the comparability of item statistics obtained under different conditions of item arrangement.

Finally, the requirement for a large final item pool for adaptive testing necessitated norming a large number of vocabulary items. A number of practical considerations militated against norming hundreds of test items on the same group of examinees, whether in a single large test or several shorter ones. Consequently, four separate tests were normed on four independent groups of persons at different times. Although all groups were nominally from the same population (college students) this raises the question of whether there were differences in group ability from one norming test to another which would detract from the direct comparability of results from different groups. Such differences, if significant, would necessitate statistical adjustments of obtained item parameters to make them comparable

across groups. Gulliksen (1950, p. 367) discusses such a statistical correction.

Thus, four norming tests were employed. Each test was administered to a different group of examinees. Mode of test administration took three different forms: conventional printed page, printed card (one item per card), and computer administration via cathode ray terminal (CRT). Items were arranged in either fixed order or random order.

The four tests were given at different times over a three-year period. Each test item was identified by a unique label, called its "item reference number". There was essentially no item overlap among the first three tests, except for a small nucleus of common items intended to permit comparisons of ability among the different examinee groups. The fourth test consisted of a number of items revised after the first three tests, and some previously administered items, including a different nucleus of common items designed to permit inter-group comparisons of ability level. Table 1 summarizes the four norming tests with respect to test medium (mode of administration), item order, test composition, number of items, and number of examinees.

The possible presence of differential effects on test difficulty due to different conditions of testing or to differences in ability among testee groups was investigated using hypothesis testing procedures. In each case a null hypothesis of no significant differences in mean test score between groups or conditions was postulated. Three null hypotheses were tested:

hypothesis 1: that variation of test medium (paper-and-pencil vs. CRT) produced no significant difference in group mean test score;

hypothesis 2: that fixed vs. random item order produced no significant difference in group mean test score;

hypothesis 3: that the four test groups did not differ significantly as to level on the ability underlying the tests.

The methods by which these three null hypotheses were tested, and the results of those tests, are detailed below.

Method

Subjects

All the subjects were undergraduate students enrolled in several university psychology courses. (This same population

Table 1

Conditions of administration, sample size and item composition for each of four norming test administrations

Test	Medium	Item Order	Number of Subjects	Number of Items				Item Reference Numbers	Calibration Subtest
				Total =	New +	Calibration +	Old		
1	Printed Card Printed Card CRT	Fixed Random Random	50 67 75	240	240	*	None	1-240	I(17 items) and II(18 items)
2	Printed Page	Random	83	180	163	17	None	241-403	I(17 items)
3	CRT	Random	81	142	125	17	None	501-625	I(17 items)
4	CRT	Fixed	144	144	47	18	79	626-672	II(18 items)
				Total	575 Unique Items				

*Note--Calibration subtests I and II were selected from among the 240 unique items of Test 1.

is the source of subjects for the studies evaluating strategies of computer-administered testing.) Participation in the norming studies was voluntary, with supplemental course points offered as an incentive in most cases.

Instruments

Item format. Each norming test was composed of 5-alternative multiple-choice vocabulary items. Each item consisted of a single word (the "stem" word) on a line, below which and indented was a column of five numbered or lettered response alternatives. Each response alternative consisted of single words or short phrases. A sample item follows:

ailment

1. accusation
2. pay
3. remedy
4. sickness
5. oil

Conditions of administration. The different subgroups of vocabulary items were administered as norming tests under three different general formats (see Table 1). These formats were designed to permit an analysis of order effects in the item norming process, and effects of computer-administration vs. paper-and-pencil norming. The conditions of test administration were:

1. 240 items printed on blank cards. Under this format, each item was printed on a separate blank computer data card. Item number was punched on each card. These cards were assembled into individual decks of 240 items each, and the decks were presented to the subjects. Subjects were instructed to circle the number of the alternative which came closest in meaning to that of the stem word. Subject responses were then punched directly into each card maintaining the order of administration. Fifty subjects were tested using card decks arranged in random item order; 67 subjects answered the same items arranged sequentially by the arbitrary item reference number. No time limits were imposed.

2. 180-item random page order printed tests. Under this format, four items were printed on a single page, and the 45 pages were assembled in random order into test booklets. Eighty-three subjects were tested in this way, using instructions similar to those used with the cards. A one-hour time limit was imposed, but was unnecessary, since every examinee finished all the items.

3. CRT-displayed items. Under this condition, items were presented to the subjects on a CRT (cathode ray tube)

computer terminal. CRT administration was employed in three different test administrations.

In test 1, the initial norming study, 75 subjects completed a random-order 240-item test (the same 240 items used in the printed cards test) on an IBM 1130 display console; to indicate their responses these subjects pressed a light-pen against the position on the screen at which the chosen alternative appeared. A 50-minute time limit was imposed.

Test 3 employed a CRT display to administer 142 items to 81 subjects; administration was controlled by a Control Data 3200 computer. Subjects responded by pressing a typewriter key bearing a number corresponding to the chosen alternative, then pressing a "send" key. A 45-minute time limit was imposed.

Test 4 was administered under conditions very similar to test 3. 144 subjects were tested for 45 minutes each at a CRT terminal of a Control Data 6400 computer. Only those items encountered by 75 or more testees were included in the analysis. 144 items met this arbitrary criterion.

Item sources and item construction. A total of 575 unique test items were normed in four major test administrations. Each item was assigned an identification number, its "item reference number". Following is a breakdown of the items normed in each administration and the sources of those items:

Administration 1 (items 1-240). Items 1 through 240 were drawn from old vocabulary tests on the assumption that the items would no longer be in general use.

--Items 1-60 were drawn from form S of the Cooperative English Test (Cooperative Test Service, 1942).
--Items 61-120 were drawn from form T of the Cooperative English Test C1 (Cooperative Test Service, 1943).
--Items 121-180 were drawn from form T of the Cooperative English Test C2 (Cooperative Test Service, 1943).
--Items 181-240 were drawn largely from the ACE Scholastic Aptitude Test, grade 11 (American Council on Education, 1947, and from the ACE Scholastic Aptitude Test, grade 11 (American Council on Education, 1952). In many cases these items were augmented with a fifth response alternative, written by project item writers and placed in sequence by a randomization procedure. The original stem and response alternatives were reviewed for contemporary relevance and revised where appropriate.

Administrations 2 (items 241-403) and 3 (items 501-625). Items 241-403 and items 501-625 were written by project item writers.[1]

Administration 4 (items 626-672). Items 626-672 were written by project item writers, in an attempt to improve 47 items normed in administrations 1 through 3 whose parameters made them unsatisfactory for use in the final item pool.

Experimental Variables

Three independent variables were of interest in these preliminary analyses of the four norming tests. These were 1) test medium, 2) order of item presentation, and 3) group ability. In each case the dependent variable was group mean test score, which was interpreted as the average within-test item difficulty.

Test medium was represented by three conditions: printed card, printed page, and CRT display. Of primary interest were any differences between paper-and-pencil administration (as implemented in the two different kinds of printed-test conditions) and CRT administration. Order of item presentation was either fixed order or random order.

In the analysis of test medium and item order effects, group ability was assumed not to vary over the four test administrations, since all four groups were drawn from the population of psychology students. That assumption needed testing, however, in light of the two-year time span between administration of the first and fourth tests.

Data Collection

Response data on the 575 vocabulary items were collected in the four major test administrations summarized in Table 1. Each person's responses to every item were scored as correct, incorrect, or not attempted. The raw data and scored data from each test administration were recorded on magnetic tape. Table 2 summarizes the specific information available from the administration of each of the four norming tests.

---

[1] Much of the early item selection, revision and item writing was done by Dennis L. Gibson and Louis J. DeWitt in 1967 and 1968. At that time this research was supported by grants from the General Research Fund of the Graduate School of the University of Minnesota, whose support is gratefully acknowledged.

Table 2

Data recorded for each subject within the four norming test administrations

	Test administration			
	1	2	3	4
testee identification number	X	X	X	X
test length	X	X	X	X
number of items attempted	X	X	X	X
number of items correct	X	X	X	X
order of item presentation	X	...	X	X
actual item responses	X	X	X	X
item scores	X	X	X	X
response latency	*	...	X	X

*available for 75 CRT testees only

Data Analysis

Hypothesis 1 postulated no difference in test performance due to effects of varying test medium. Hypothesis 2 postulated no difference due to the effects of fixed vs. random item order. Hypothesis 3 postulated no differences in word knowledge ability among the four groups of psychology students tested. To determine whether any of the three independent variables affected item parameter estimates, a series of analyses was performed. The order in which these analyses was conducted paralleled the chronological sequence of test administrations 1 through 4, and the results of each analysis were used as a basis for the designs of the data collection in successive administrations.

Test medium and item order effects. In order to test hypotheses 1 and 2, some basis for comparison of test difficulty obtained under different conditions of test medium and item order was required. This basis was provided in the design of the data collection for test 1.

"Accuracy score", the subject's proportion correct among those items attempted[2], was employed as the dependent variable in a one-way analysis of variance. This proportion, rather than number correct, has the advantage of eliminating from each total score any speed effects which might have been operating in spite of the essentially non-speeded nature of the tests. Furthermore, each individual's accuracy score may be interpreted as the intra-individual average difficulty of the items he/she attempted. Comparisons of mean accuracy scores among treatment groups was considered to be a more stable indicator of differential treatment effects than comparison of inter-group differences in the parameters of items singly.

The administration of test 1 involved 192 testees randomly assigned to take the 240-item test under one of three conditions: 1) paper-and-pencil with fixed item order; 2) paper-and-pencil with random item order; and 3) CRT administration with random item order. Table 3 details those three conditions and the number of subjects tested under each one. Subjects were randomly assigned to conditions; each of the three experimental conditions represents a fixed treatment characterized by its unique combination of test medium and item order. Thus a 1 x 3 fixed effects analysis of variance was used to test hypotheses 1 and 2 simultaneously. Post-hoc contrasts were planned to identify significant sources of variation if the anova produced significant results.

[2]Throughout this report, the term "encounter" will be used to denote a presentation of a test item to a testee. An item is "encountered" by the testee if he/she has had an opportunity to read and respond to it. An item is "attempted" if the testee both encounters it and responds. Thus, an item not encountered is not attempted.

Table 3

Distribution by test condition of the 192 examinees administered Test 1

	Test Medium	
Item Order	Printed Card	CRT
Fixed order	Subgroup #1 N = 67	
Random order	Subgroup #2 N = 50	Subgroup #3 N = 75

Table 4

Data available for comparison of group ability across norming tests 1, 2, 3 and 4

	Test Administration			
Test Score	1	2	3	4
Calibration Subtest I (17 items)	N=192	N=83	N=81	
Calibration Subtest II (18 items)	N=192			N=136*

*Eight individuals in test administration group 4 did not complete Calibration Subtest II.

Group Differences in ability. In order to provide a basis for comparison of average ability among groups, a nucleus of 17 common items was selected from among the 240 items of test 1.[3] This nucleus was subsequently incorporated within tests 2 and 3. Such a nuclear subtest is useful for correcting item analysis results for differences in ability from one norm group to another (Gulliksen, 1950, p. 367). This 17-item common subtest will be referred to hereafter as calibration subtest I. In test 4, the calibration subtest was substantially revised, and the resulting 18-item subset, calibration subtest II, was developed as the basis for future inter-group comparisons. Table 4 summarizes the calibration subtest data collected and available for comparisons of ability among the groups administered tests 1, 2, 3, and 4 respectively.

From Table 4 it can be seen that comparable performance data on calibration subtest I were available for groups 1, 2, and 3. A similar comparison of performance on calibration subtest II was possible for groups 1 and 4 only.

Prior to performing inter-group comparisons of the four groups on the two calibration subtests, each of the subtests was subjected to an analysis of its own psychometric properties, including subtest mean, variance, and Kuder-Richardson formula 20 internal consistency reliability for each group, as well as a traditional item analysis. (It was this analysis of calibration subtest I which led to its replacement by calibration subtest II.)

For groups 1, 2, and 3, hypothesis 3 (that of no differences in ability among groups) was tested using a 1 x 3 fixed effects analysis of variance, with accuracy score on calibration subtest I as the dependent variable. Post-hoc contrasts of all intergroup pairs were planned if the null hypothesis was rejected.

For test groups 1 and 4, a t-test contrast of mean accuracy scores on calibration subtest II was performed.

Results

Test Medium and Item Order Effects

The test 1 data for the three test medium/item order conditions are summarized in Table 5. The summary data include grand mean and variance, and within-group means, variances and sample sizes for the three conditions. Table 6

---

[3] Construction of the 17-item subset was accomplished by Dennis L. Gibson in 1969.

Table 5

Test 1 group means and variances under each of three conditions of administration

Test Administration Condition	N	Accuracy Score	
		Mean	Variance
1. fixed item order on printed cards	67	.71	.009
2. random item order on printed cards	50	.72	.009
3. random item order on CRT	75	.69	.013

Table 6

Analysis of variance results for testing the hypothesis of no difference in Test 1 performance due to varying conditions of administration

Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	F
Between conditions	2	.0205	.0103	.990
Within conditions	189	1.9593	.0104	
Total	191	1.9798		

Table 7

Item difficulty (p) discrimination (rbis) and KR-20 reliability coefficients (r_{xx}) for calibration subtests I and II within each norming administration

Calibration Subtest I							Calibration Subtest II				
	Group 1		Group 2		Group 3			Group 1		Group 4	
Item	p	rbis	p	rbis	p	rbis	Item	p	rbis	p	rbis
20	.984	.341	.988	.481	.955	.692	52	.409	.554	.610	.466
54	.577	.288	.512	.427	.590	.369	59	.466	.561	.463	.509
64	.990	.948	.988	-.215	.986	-.255	85	.728	.648	.765	.544
71	.989	.947	.976	-.254	.986	.368	88	.838	.548	.809	.512
77	.942	.381	.878	.581	.893	.540	96	.948	.927	.890	.508
78	.979	.381	.976	.253	.959	.241	99	.984	1.000	.993	.397
88	.838	.548	.841	.667	.824	.645	101	.890	.814	.882	.558
96	.948	.927	.904	.698	.933	.458	113	.426	.580	.485	.434
99	.984	1.000	1.000	0.000	.945	.500	114	.351	.600	.338	.634
116	.568	.402	.494	.277	.667	.418	116	.568	.402	.537	.288
126	.953	.660	.988	-.085	.973	.013	152	.220	.451	.296	.530
152	.220	.451	.265	.496	.292	.309	162	.316	.493	.267	.413
159	.337	.337	.277	.583	.310	.362	166	.120	.611	.163	.438
172	.453	.252	.458	.418	.500	.316	173	.775	.655	.858	.538
183	.681	.447	.683	.540	.622	.717	180	.199	.526	.216	.212
201	.848	.273	.866	.301	.817	.383	183	.681	.447	.694	.609
230	.344	.163	.434	.175	.397	.038	188	.634	.645	.575	.491
							237	.094	.324	.104	.374
r_{xx}	.52		.56		.80			.76		.78	

contains the analysis of variance table for the combined tests of hypotheses 1 and 2. No statistically significant effects due to treatments were found, based on the obtained F-ratio of .990.

Group Ability Effects

Analysis of calibration tests. Table 7 contains summary data for the psychometric analyses of calibration subtest I within tests 1, 2, and 3, and the analysis of calibration subtest II within tests 1 and 4. The item analysis data for the three administrations of calibration subtest I reveals that that subtest consisted primarily of very easy items, some of which had low discriminations (biserial correlations with total score, rbis), and that its internal consistency reliability was low in two of the three groups. Calibration subtest II, on the other hand, employed a rectangular distribution of item difficulty, consisted of items with moderate to high discriminations, and was consistently more reliable than subtest I. The unreliability of calibration subtest I necessitated caution in interpreting the comparison of ability among groups 1, 2, and 3.

Ability level comparison. Table 8 contains summary data--means, standard deviations and sample sizes--for the four test groups on the two calibration subtests. Table 9 contains the analysis of variance table for the test of significant differences among groups 1, 2, and 3 in their performance on calibration subtest I. The obtained F-ratio of .16 was not statistically significant.

Table 10 contains the results of the pooled variance t-test of differences between mean scores on calibration subtest II. The obtained t-ratio of .53 was not statistically significant.

On the basis of these results, the four test administration groups were considered not to differ from one another on the ability underlying the two calibration subtests.

Not only were there no significant mean differences in ability among the four groups, but also, within each subtest, the item difficulty parameters were very similar across groups. Table 7 shows that on calibration subtest I, the largest difference in item difficulty (p) between groups was .173 (item 116 for groups 2 and 3); on calibration subtest II the largest p-value difference was .110. Within each subtest, the correlation between item p-values from one group to another exceeded .98 for every pair of groups.

Table 8

Mean and standard deviation of scores on calibration subtests I and II within each of the four norming test administrations

Subtest	Test Administration 1	2	3	4
Calibration Subtest I				
Mean	.742	.736	.747	...
Standard deviation	.114	.124	.136	...
N	192	83	81	...
Calibration Subtest II				
Mean	.542	...	...	.553
Standard deviation	.181	...	...	.188
N	192	...	...	136

Table 9

Analysis of variance table for testing the significance of differences in group mean performance on Calibration Subtest I

Source of Variance	Degrees of Freedom	Sums of Squares	Mean Square	F
Between groups	2	.0047	.0023	
Within groups	354	5.1956	.0147	.16
Total	356	5.2003		

Table 10

Means and standard deviations on Calibration Subtest II for groups 1 and 4, and the t-test of the difference in means

Group	Mean	Variance	t
1	.542	.0328	
			.53
4	.553	.0355	

Discussion

The statistical analyses concerned with differential effects on test 1 performance due to varying the conditions of testing resulted in non-significant differences. On the basis of the data reported, the decision was made to treat test data obtained under any of the three conditions as equivalent. On that basis, item response data obtained by paper-and-pencil testing within test 2 was considered comparable to item response data gathered by CRT testing in tests 3 and 4.

The investigation of differences in group ability across the four tests also found no significant differences. Therefore, the four test groups were considered equal in overall ability on the dimension relevant to this testing research, and no adjustments for group ability were made to the item parameters obtained within any of the four groups.

Another encouraging result of comparing the four test groups was the observed stability of the calibration subtest item difficulty parameters from one group to another. The high correlations (.98 or higher) between the item difficulty values across groups has important implications for the stability of the parameters of the remaining norming test items whose difficulty in different groups could not be observed.

Another result of the inter-group comparisons that merits discussion is the low internal consistency of calibration subtest I in two out of the three test groups. As shown in Table 7, that subtest had Kuder-Richardson formula 20 internal consistency reliabilities of .52, .56 and .80 in test groups 1, 2, and 3 respectively. With the exception of the group 3 reliability, these values were much lower than is ordinarily considered satisfactory in ability testing, a fact which makes calibration subtest I a dubious standard for comparing groups. The unreliability of subtest I was probably due to the distribution of the difficulty values of its items, almost half of which had p-values of .90 or larger (Table 7). Such a distribution would tend to reduce the variance in total scores, and thereby reduce reliability. The effect of these observations is that calibration subtest I is a rather poor basis on which to compare groups. Calibration subtest II was constructed specifically to remedy the psychometric deficiencies of subtest I. Subtest II employed a rectangular distribution of items across the difficulty continuum and, as Table 7 shows, was consistently more reliable in comparison to subtest I. Calibration subtest II was incorporated into test 4, and is intended for incorporation in future norming tests.

DEVELOPMENT OF THE FINAL ITEM POOL

Rationale

Since there were no mode of administration or item order effects in the preliminary analysis, and no inter-group ability differences, the item response data from tests 1 through 4 were not adjusted to account for differences in conditions of administration or among examinee groups. Once this decision was made, the only remaining independent variables of interest were the items themselves. Difficulty and discrimination parameters for each item were then computed as the next step in structuring the item pool for use in adaptive testing.

Two sets of parameters were estimated for each item. The first set consisted of traditional item analysis parameters--difficulty (proportion correct) and discrimination (biserial correlation between item score and total score). The second set of item parameters were those of the two-parameter normal ogive item characteristic curve for each item. The traditional item parameters were to be used for selecting the final item pool. The normal ogive parameters have properties which make them particularly suitable for use in adaptive testing and for "item banking". Both these points will be discussed below.

For the purpose of item analysis, each examinee's total score was expressed as the proportion correct among the items he/she attempted; this is the "accuracy score" mentioned earlier. The traditional item difficulty parameter, proportion correct, was defined as the proportion of the examinees encountering an item who answered it correctly. Both accuracy score and the item difficulty index, then, minimize the effect of speed on the item analysis results.

The biserial correlation between item score and accuracy score was employed as the traditional item discrimination index. The biserial, rather than the point-biserial correlation, was selected because of the tendency of the latter statistic to be biased in favor of items of median difficulty (Gulliksen, 1950, p. 393). Since the purpose of item norming was to develop a large pool of items whose difficulty values were rectangularly distributed over the useful range of difficulty, use of a statistic which favors the median would have distorted the characteristics of the final item pool.

A total of 575 unique vocabulary test items were normed. Each item was included in one or more of four separate norming tests administered to samples selected from a population

of undergraduate psychology students. Since the best estimate of a population parameter is obtained by using all the data available, item parameters of items which occurred in more than one test were estimated by pooling the data over the two or more tests in which any item was repeated.

Normal ogive item parameters were determined for all items as the final step of item norming. The normal ogive model item parameters, "a" and "b", are analogous to the discrimination and difficulty indices of the traditional model, but have the advantage of being invariant from group to group, granted the assumptions of the model (Lord & Novick, 1968, pp. 374-379). The difficulty index, "b", is expressed in the same metric as that on which ability is scaled under the model. It corresponds to the level of ability (measured continuously) at which an individual has .50 probability of responding correctly to the item. The discrimination index, "a", is proportional to the slope of the normal ogive item characteristic curve at "b"; the higher the slope at that point, the more discriminating the item. Details of the computation of these parameters are given below.

Estimates of the normal ogive parameters were desirable for several reasons. First, the normal ogive item difficulty parameter, "b", is a function of both proportion correct and the biserial correlation of item score and ability; thus item scaling in terms of "b" is based on more information than is contained in the traditional difficulty index. This results in more precise ordering of items by difficulty, which may be a critical issue in evaluating the relative merits of adaptive tests (Urry, 1970). Second, some adaptive test strategies require knowledge of latent trait model item parameters in order to estimate an examinee's ability (e.g., Owen, 1969; Urry, 1970). The normal ogive model is probably the most widely known of the latent trait models presently formulated. Third, the invariance properties of the parameters of the normal ogive item characteristic curve permit the transformation of parameters obtained within a specific subgroup of the population into parameters relative to any other subgroup (Lord & Novick, 1968, p. 381). This last point has important implications for the possibility of establishing large "banks" of items of known parameters, and employing those items to estimate an examinee's ability relative to the ability distribution within any subgroup of a well-defined population.

Once the two sets of item parameters had been estimated, items were selected for the final item pool. Since a large

item pool containing items at all levels of difficulty was desired, the item discrimination index was the critical parameter for inclusion of any item within the final item pool.

Method

Item Parameter Estimation

Traditional item parameters. The estimates--p_i, r_{it}, and r'_{it}--of the traditional item parameters π_i, ρ_{it}, and ρ'_{it}, were obtained using an item analysis program for binary-scored (right-wrong) items. The item parameter estimates were calculated by the following methods:

$$p_i = \hat{\pi}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} u_{ij} \qquad [1]$$

where p_i = proportion correct for item i

N_i = number of testees who attempted item i

u_{ij} = 1 for a correct response to item i by testee j
0 for an incorrect response

$$r_{it} = \hat{\rho}_{it} = \frac{(\bar{X}_i^+ - \bar{X}_i^-)}{s_x} (p_i q_i)^{\frac{1}{2}} \qquad [2]$$

$$r'_{it} = \hat{\rho}'_{it} = \frac{(\bar{X}_i^+ - \bar{X}_i^-)\ p_i q_i}{s_x\ z_p} \qquad [3]$$

where r_{it} = point-biserial item-test correlation

r'_{it} = biserial item-test correlation

p_i is as defined for formula 1

q_i = 1-p_i

$\bar{X}_i^+$ = mean test score of testees answering item i correctly

$\bar{X}_i^-$ = mean test score of testees answering item i incorrectly

s_x = standard deviation of total test scores

z_p = ordinate of the unit normal curve at p_i

The test score used was "accuracy score", the proportion correct among the items actually encountered by an examinee. An item encountered but not attempted was scored as incorrect; an item not encountered (i.e., not reached) was not scored. Accuracy score for each individual (X_j) was computed by formula:

$$X_j = \frac{1}{k_j} \sum_{i=1}^{k_j} u_{ij} \qquad [4]$$

where u_{ij} is defined as for formula 1 and k_j is the number of items encountered by testee j

Separate traditional item parameter estimates were obtained for each item on each test administered. On the assumption that all testees were sampled from the same population, data were pooled across tests for those items common to more than one test. On this basis weighted estimates of the parameter P_i were generated, with the weights proportional to the number of subjects within each test. The formula used was:

$$\bar{P}_i = \frac{1}{N_{i.}} \sum_{j=1}^{t} N_{ij} P_{ij} \qquad [5]$$

where $\bar{P}_i$ = the pooled estimate of item difficulty

N_{ij} = number of persons encountering item i within test j

P_{ij} = proportion correct for item i within test j

$N_{i.} = \sum_{j=1}^{t} N_{ij}$

and t = the number of tests in which the item appeared

Pooled-data estimates of the point-biserial item-ability correlation were obtained by a similar weighted-average method.

Normal ogive item parameters. Two parameters, "b" and "a", were of interest for the normal ogive item characteristic curve for each item. The difficulty parameter, "b", corresponds to the point on the ability continuum at which the probability of a correct response to a given item is exactly .50; the discrimination parameter, "a", is the reciprocal of the standard deviation of the item characteristic curve, and is proportional to the slope of the curve at "b". Both "a" and "b" may be expressed as functions of the biserial correlation of item score and ability (Lord, 1952; Tucker, 1946). Assuming a normal distribution of ability within the examinee group, and a metric chosen such that the mean ability is equal to 0.0, with standard deviation equal to 1.0, the following relationships hold (e.g., Baker, 1969):

$$p_i = \frac{1}{\sqrt{2\pi}} \int_{y}^{\infty} e^{\frac{-z^2}{2}} dz \qquad [6]$$

$$b_i = \frac{-y}{r'} \qquad [7]$$

$$a_i = \frac{r'}{(1-r'^2)^{\frac{1}{2}}} \qquad [8]$$

where p_i = the proportion correct, as in formula 1

z = a normal deviate

Y = the inverse of the cumulative normal distribution function at p_i (a normal deviate)

$r' = r'_{i0}$ = the biserial correlation of item score and ability

For the purposes of this research, the point-biserial item-test correlation (r_{it}) obtained from the traditional item analyses were transformed to estimate the value of the biserial item-ability correlations (r'_{i0}). The transformation used was:

$$r'_i = r_{it}(p_i q_i)^{\frac{1}{2}} (F(Y_i))^{-1} \qquad [9]$$

where r_i, r'_i, p_i, and q_i are as previously defined

$F(x)$ is the unit normal density function

Y_i is as defined above for formulas 6 and 7

This transformation, in somewhat different notation, is given in Lord and Novick (1968, p. 340).

The parameters a_i and b_i were estimated from the biserial correlations thus obtained, using equations 7 and 8. The value of a_i is theoretically unrestricted; however, as the value of r'_i increases beyond an absolute value of .90, a_i approaches infinity rapidly. To prevent this, a_i was arbitrarily set to 3.00 whenever the absolute value of r' was greater than or equal to .90.

Selection of the Final Item Pool

The purpose of the norming efforts reported here was to select, from among the developmental item pool of 575 items, a final item pool of vocabulary items whose psychometric characteristics make them suitable for inclusion in adaptive tests for use within the population sampled. An adaptive test may require a number of items at every level of difficulty. For that reason, a relatively large final item pool of highly discriminating items, rectangularly distributed across the useful range of item difficulty, was desired.

Item selection was accomplished on the basis of the traditional item parameters rather than the normal ogive parameters, since the normal ogive item parameters were estimated by approximate methods. The item discrimination index was the critical parameter for selection of items for the final item pool. An arbitrary criterion of .295 was established as the cutting point; only those items with a biserial item-total correlation exceeding .295 were selected for inclusion in the final item pool. This criterion was established for two reasons. First, it is generally representative of the lower limit of discrimination acceptable for construction of traditional ability tests. Second, it is statistically significant at the .10 level for the smallest number of subjects responding to any item normed i.e., even where as few as 75 testees responded to an item, .295 was significantly different from a zero correlation at $p \leq .10$, using a significance test suggested by Alf and Abrahams (1971).

Results

Traditional Item Parameters

Appendix A contains a listing of the traditional item parameters obtained for each of the 575 items; these are the pooled data estimates for those items which were repeated in more than one test. The composition of the item pool is summarized in Table 11, which lists the obtained traditional parameters in a cross-tabulated count of item discrimination values within difficulty intervals. The number within each cell of Table 11 indicates the number of items whose parameters fell within the discrimination and difficulty interval defined by that cell. The total number of items at any difficulty or discrimination level is indicated by the corresponding marginal total.

The range of item difficulty observed was from a proportion correct of .029 to 1.000. Item discrimination values (biserial item-total correlations) ranged from -.37 to 1.00. The majority of selected items had discrimination indices in the moderate range (.30 to .60), with the highly discriminating items tending also to be very easy ones. The latter observation may reflect sampling error, but in general there is a moderate interaction between the difficulty and discrimination parameters, which covary positively. That is, the more discriminating items tend to be relatively easy; the more difficult items tend to be less discriminating.

Table 11 reveals a need for more items in the moderate to very difficult range. Nonetheless, there are sufficient numbers of items at all difficulty levels for use in selecting the final item pool.

Normal Ogive Item Parameters

The estimated parameters of the normal ogive model item characteristic curves are also listed in Appendix A. Table 12, a cross-tabulation by discrimination (a) and difficulty (b) indices, summarizes the distribution of these estimates. (For the purposes of the present research, extreme values of the difficulty parameter ($|b| \geq 3.60$) are not distinguished from one another.)

The normal ogive model ability parameter, θ, can be interpreted similarly to a normalized standard score (z-score); thus virtually the entire population have ability θ ranging from -3.00 to +3.00. 403 items fell within this range, 400 of which had discrimination parameters, "a", larger than 0.0.

Table 11

Number of items at each level of difficulty and discrimination, based on traditional item parameters, p and r_{it}

Discrimination (r_{it})	Proportion Correct (p) 0 to .099	.10 to .199	.20 to .299	.30 to .399	.40 to .499	.50 to .599	.60 to .699	.70 to .799	.80 to .899	.90 to 1.00	Total
.90 - 1.00										7	7
.80 - .899									2	8	10
.70 - .799				2	1	3		1	6	11	24
.60 - .699	1	1	1	2	9	5	8	6	10	9	52
.50 - .599		3	7	3	6	11	18	8	7	15	78
.40 - .499	1	6	12	19	8	16	12	14	8	10	106
.30 - .399	3	10	15	15	6	12	7	8	6	10	92
.20 - .299	1	7	3	15	15	9	4	2	5	10	71
.10 - .199		15	8	12	5	2	2	1	1	1	47
0 - .099	1	11	13	13	4	3	1	2		2	50
$r_{it} < 0$	4	15	5	2					1	3	30
Total	11	68	64	83	54	61	52	42	46	86	567*

*Eight of the original 575 items were dropped from consideration due to irregularities of administration.

Table 12

Number of items at each level of difficulty (b) and discrimination (a), for the normal ogive item parameters

Discrimination a	Difficulty (b) b> 3.6	2.8- 3.59	2.0- 2.79	1.2- 1.99	.4- 1.19	-.4- .39	-1.2- -.39	-2.0 -1.19	-2.8 -1.99	-3.6 -2.79	b < -3.6	Total
a > 2.00									7			7
1.00 - 1.999						3	1	12	9	3		28
.90 - .999					3	4	2	5	3	4		21
.80 - .899		1	1			6	6	3	1			18
.70 - .799					4	6	9	7	2	2		30
.60 - .699			1	4	3	11	19	4	6	3	4	55
.50 - .599		1	1	7	12	11	15	7	6	3	2	65
.40 - .499		3	11	10	13	14	7	6	8	3	5	80
.30 - .399	4	4	8	13	10	8	8	7	1	3	8	74
.20 - .299	7	2	2	10	10	7	8	4		3	13	66
.10 - .199	22	2	4	6	2	2		1	1		4	44
.00 - .099	38	2		1	1	1	1				5	49
a < 0	4								2	3	21	30
Total	75	15	28	51	58	73	76	56	46	27	62	567*

*Eight of the original 575 items were dropped from consideration due to irregularities of administration.

Description of the Final Item Pool

Application of the minimum biserial criterion for item selection resulted in a final item pool of 369 items. Appendix B, a cross-tabulation of the items in the final item pool by discrimination and difficulty parameters, lists the selected items by item reference numbers (see Appendix A for specific information on each item).

Inspection of the marginal totals of the difficulty distribution in Appendix B reveals that the distribution of difficulty values of this final item pool was not rectangular as desired. Rather, there is a predominance of items in the moderate-to-easy range, with fewer very difficult items available than easy ones. Inspection of the marginal frequencies of Table 11 shows that the distribution of items by difficulty favors the relatively easy items. Far more easy items (p-values greater than .60) were normed than difficult ones (p-values less than .40); difficulty parameters of 115 items fell in the median difficulty range (p-values between .40 and .60). However, as Appendix B shows, with the exception of very difficult items ($p < .10$) the final item pool had a minimum of 20 items available in each decile of the range of item difficulties.

DIMENSIONALITY OF THE FINAL ITEM POOL

Rationale

Based on the analyses reported above, a pool of 369 items was selected for use in computer-administered ability testing research. A requirement of this final item pool is that its items measure a single dimension of ability. Not only is unidimensionality of the item pool a basic assumption underlying adaptive testing strategies (see Weiss & Betz, 1973), it is also fundamental to the appropriateness of the use of the normal ogive model for scaling items and estimating ability (Lord & Novick, 1968, p. 366). These normal ogive item parameters will be employed in some of the adaptive testing strategies examined in the course of the parent research project. Because of the importance of unidimensionality of the item pool for the uses to which it will be applied, some evaluation of that assumed unidimensionality was necessary.

Rigorous investigation of the hypothesis of unidimensionality is difficult since statistical procedures for testing such a hypothesis in the absence of distributional assumptions are not presently available. When distributional assumptions are made (such assumptions are fundamental

to use of the normal ogive as a model for the item characteristic curve) factor analytic procedures are available, but their results are not definitive, since the number of common factors extracted from a correlation matrix is partly a function of the index of relationship used. Lord and Novick (1968, p. 375) point out that multivariate normality of the variables underlying a set of dichotomous item statistics, and a single common-factor structure, are sufficient but not necessary conditions for confirming the assumption of unidimensionality of the latent variable space underlying a set of test scores.

In the absence of a definitive test of the unidimensionality assumption, unidimensionality was evaluated indirectly. At an intuitive level, at least, the items comprising the final item pool appear to be homogeneous. All are vocabulary test items; all have the same format; all require the same task of the examinee; the only apparent difference in the items is in the specific knowledge required in order to respond correctly to each. Granted that the final item pool seemed homogeneous, however, the question was investigated further by making a set of predictions about its expected characteristics if the unidimensional case obtained, then testing those predictions with the norming data.

If the item pool were unidimensional, it would be expected that the matrix of tetrachoric interitem correlations would have just one common factor. Demonstration of that prediction in the data would constitute sufficient evidence of unidimensionality, as mentioned above. For several reasons, however, it was not possible to study the dimensionality of the intercorrelations of the 369 items in the final item pool. First, the sheer size of the correlation matrix (369 by 369) involved was too large for available computer facilities. Secondly, however, the research used in the norming studies did not yield data amenable to such an analysis. No testee in the norming studies had completed more than 240 items out of the 375 originally studied. Furthermore, no item was completed by more than 192 subjects within any one norming test. Therefore, the subjects-to-variables ratio in the analysis would have been unfavorable to a meaningful factor analysis, with almost twice as many variables as subjects.

To circumvent these problems, yet to obtain some evidence about the unidimensionality of the final item pool, an alternative procedure was followed. Random subsets of items were sampled from the 369-item final pool. For each random subset, the interitem correlation matrix was computed, and factor analysis was performed on each such matrix. In addition to these analyses, factor analyses were performed on several similar matrices constructed from computer-

generated random item response data. Confirmatory evidence of unidimensionality in the real item pool would be 1) that the first common factor in each real data matrix would be a general factor which would account for a major proportion of the common variance and on which all variables would have significant loadings; 2) that the second and subsequent factors extracted from each real data matrix would account for much smaller, and approximately equal proportions of common variance; 3) that the loadings of the real variables (test items) on the first common factor would reflect a positive manifold in the test items, i.e., that they would load uniformly in one direction; and 4) that none of the matrices of random generated data would satisfy criteria 1, 2, and 3. These first three criteria are similar to those applied in a related study, without item sampling, by Indow and Samejima (1966). The fourth criterion, which involves contrasting the factor structure and factor contributions of real test data with those of random data, was intended as a safeguard against random sampling fluctuations.

Further confirmatory evidence for unidimensionality also involves the first factor loadings of the items sampled for each factor analysis. Henrysson (1962) has shown that each first factor loading closely approximates the biserial correlation between item score and the total score, when a complete set of items is factor-analyzed. In a set of unidimensional test items, the first factor may be thought of as the "latent trait," inperfectly measured. Then the loading of each item on the first factor is an estimate of the item "reliability," its saturation with the latent trait. Another estimate of the same quantity is provided by the traditional item discrimination index, which is the correlation between item score and total score on the test. In the unidimensional case, the test itself is highly saturated with the latent trait; therefore each item-test correlation should correspond very closely to the item's loading on the first factor. The strength of such a correspondence has implications for the tenability of the assumption of unidimensionality.

Such an analysis was conducted for each of the randomly selected subsets of items mentioned above. Factor analyses were performed using the matrices of tetrachoric interitem correlations. For each item, its first factor loading was compared with its item-test biserial correlation obtained from the analysis of the norming tests. The expectation was that the two indices would correspond highly, lending support to the assumption of unidimensionality.

Method

The method used to evaluate the unidimensionality hypothesis involved contrasting the results of factor analyses

of empirically obtained interitem correlation matrices against the results of similar analyses on intercorrelations of computer-generated random binary variables. Assuming unidimensionality of the items of the final item pool, certain specific predictions were made about the structure underlying any matrix of intercorrelations among samples of its items. The degree to which such predictions were confirmed in the empirical responses of examinees to the items of the final item pool was taken as a measure of the tenability of the assumption of unidimensionality. Additional confidence in that assumption would be justified if it could be shown that such predictions, if verified in the empirical data, do not hold in random data, i.e., in sets of random item responses in which a single dimension of ability is known not to underly item performance.

In order to employ factor analysis meaningfully, a minimum subjects-to-items ratio of about 10-to-1 was adopted. Since the largest norming test (test 1) had 192 subjects, a subtest drawn therefrom should be about 20 items in length to approximate the desired ratio. Accordingly, from among the test 1 items selected for the final item pool, 20 items were sampled by a randomization procedure to constitute a subtest. This random item-sampling procedure was repeated, without replacement, until six independent 20-item subtests were constructed from the items of test 1. Then, from the original scored item responses of test 1, the six data sets were drawn, each consisting of the responses of the 192 subjects to a 20-item subtest. Corresponding to these, six matrices of random 1's and 0's (i.e., random "item" scores) were computer-generated, each matrix consisting of the responses of 192 "subjects" to 20 hypothetical "items."

For each of the twelve subsets of item scores, six of which contained real data and six containing random data, a 20-by-20 matrix of interitem correlations was constructed. These matrices were obtained using program TETREST (McBride, 1974), a computer program which analyzes scored test data and yields matrices of interitem tetrachoric correlation coefficients. The tetrachoric r was selected as the index of interitem relationship because of its invariance under differences in the marginal proportions of pairs of items; the fourfold point correlation (phi) is well known to be sensitive to such differences (Carroll, 1961).

Each of the twelve intercorrelation matrices was analyzed by an unrotated principal axes factor analysis, with communalities estimated by the absolute value of the largest correlation in each row of the matrix. The first nine factors were extracted from each matrix.

The criterion for accepting the hypothesis of a unitary attribute underlying the final item pool was satisfaction of the following predictions about the results of factor analyses of the real and random data.

1. Each of the real data matrices was expected to show the same factor structure comprised of a single large general factor. The random data matrices were not expected to show such a general factor.

2. After extraction of the large first factor, the second and subsequent factors in the real data were expected to have eigenvalues of approximately the magnitude of spurious factors due to sampling error; these were expected to be of about the same magnitude as the largest factor in the random data matrices.

3. Within the real data matrices, each of the test items was expected to load significantly and in a uniform direction on the first factor, in contrast to the random data variables, some of which should load positively and some negatively on the respective random factor. Significant loading in a uniform direction would be the case in a true single-factor structure when all the variables were preselected on the basis of positive correlations with some measure of that factor; the 120 real-data items from the final item pool were so selected. Failure to meet this requirement in any of the real data matrices would be evidence that the first common factor in that matrix was not approximately collinear with norming test score (the criterion against which the discrimination index was calculated in the item analysis phase of the present study).

4. Within each real data matrix, the first factor loadings of the items were expected to correspond closely to the item discriminations obtained from the item analyses. The item discriminations were the biserial correlations between item score and score on the longer norming test; the first factor loadings were those occurring within a 20-item subset of the longer test, analyzed as a subtest. Although the subtests were not independent of the norming tests, the high correspondence between the indices in question would not be expected to obtain uniformly across the several subtests unless the items of each subtest were all saturated with the same common factor and that factor was the first common factor contained in the norming test. The latter premise would be highly unlikely unless the norming test measured essentially a single common factor. Evaluation of the correspondence between item discriminations and their first factor loadings was accomplished by calculating product-moment correlation coefficients between the two indices for the items of each subtest. Finally, that correlation coefficient was calculated for all 120 real items combined.

Results

Factor contributions. The 12 matrices of tetrachoric interitem correlations and the principal axis factor loadings of the first nine factors extracted in each analysis are shown in Appendix C.

Table 13 lists the eigenvalues and percentage of common variance accounted for by each of the first four factors for each of the six factor analyses for both real and random data matrices. In the six real data matrices, the first factor accounted for 50% to 59% of the common variance, and the second factor accounted for 13% to 20% of the common variance. The first factor in each case, therefore, accounted for about three to four times as much common variance as the next largest factor.

In the random data, the largest factor within each subset accounted for little more common variance than the second and lesser factors. First factors of the random data accounted for about 25% to 28% of the common variance, approximately half of the proportion of common variance accounted for in the real data. Second factors of the random data accounted for more of the common variance in every case than did the second factors of the real data. Similar findings were observed for the third and fourth factors of the random data in comparison to those of the real data. Since the random data arose out of a deliberate randomization process, the logical "factor structure" underlying them can be considered to reflect sampling fluctuations around a true null relationship.

Figure 1 compares the eigenvalues of the real data matrices with those of the random data matrices, showing ranges and medians of the nine eigenvalues extracted from each of the six real and six random data factor analyses. Figure 1 shows that in the real data the first factor extracted by far the largest amount of variance, while the second factor extracted little more common variance than the later ones. Indow and Samejima (1966) interpreted strikingly similar results as arising from a one-dimensional latent space. The random data factor contributions depicted graphically in Figure 1 are notably different from the real data factor contributions in one principal respect: the random data lacked a general factor which accounted for any substantial amount of the common variance. Furthermore, none of the random factors even approximated the magnitude of the first factor from the real data analyses. The random data factor contributions are remarkably like those of the second and subsequent factors in the real data. These findings provide evidence for the single-factor hypothesis in the real data.

Table 13
Eigenvalues (factor contributions) and percent of common variance accounted for by the first four factors from each of the six factor analyses for real and random data

	Real Data				Random Data		
	Factor	Eigenvalue	% of common variance		Factor	Eigenvalue	% of common variance
Subtest 1	1	6.37	52.50	Sample 1	1	1.22	26.30
	2	1.59	13.07		2	1.02	22.06
	3	1.26	10.40		3	.92	19.79
	4	1.07	8.81		4	.76	16.34
			84.79				84.49
Subtest 2	1	5.34	50.23	Sample 2	1	1.27	25.14
	2	2.10	19.80		2	1.08	21.48
	3	1.21	11.36		3	.83	16.50
	4	1.05	9.87		4	.74	14.64
			91.26				77.76
Subtest 3	1	5.92	56.04	Sample 3	1	1.23	26.35
	2	1.51	14.32		2	1.01	21.56
	3	1.18	11.19		3	.82	17.59
	4	.96	9.07		4	.73	15.64
			90.63				81.15
Subtest 4	1	5.50	54.17	Sample 4	1	1.19	27.61
	2	1.57	15.43		2	.87	20.14
	3	1.00	9.84		3	.80	18.54
	4	.94	9.32		4	.66	15.35
			88.76				81.64
Subtest 5	1	5.84	58.67	Sample 5	1	1.25	25.90
	2	1.32	13.24		2	1.02	21.11
	3	1.09	10.96		3	.90	18.51
	4	.81	8.10		4	.78	16.11
			90.96				81.63
Subtest 6	1	4.62	50.15	Sample 6	1	1.16	24.97
	2	1.27	13.79		2	1.08	23.42
	3	1.04	11.25		3	.98	21.14
	4	.91	9.90		4	.62	13.37
			85.09				82.90

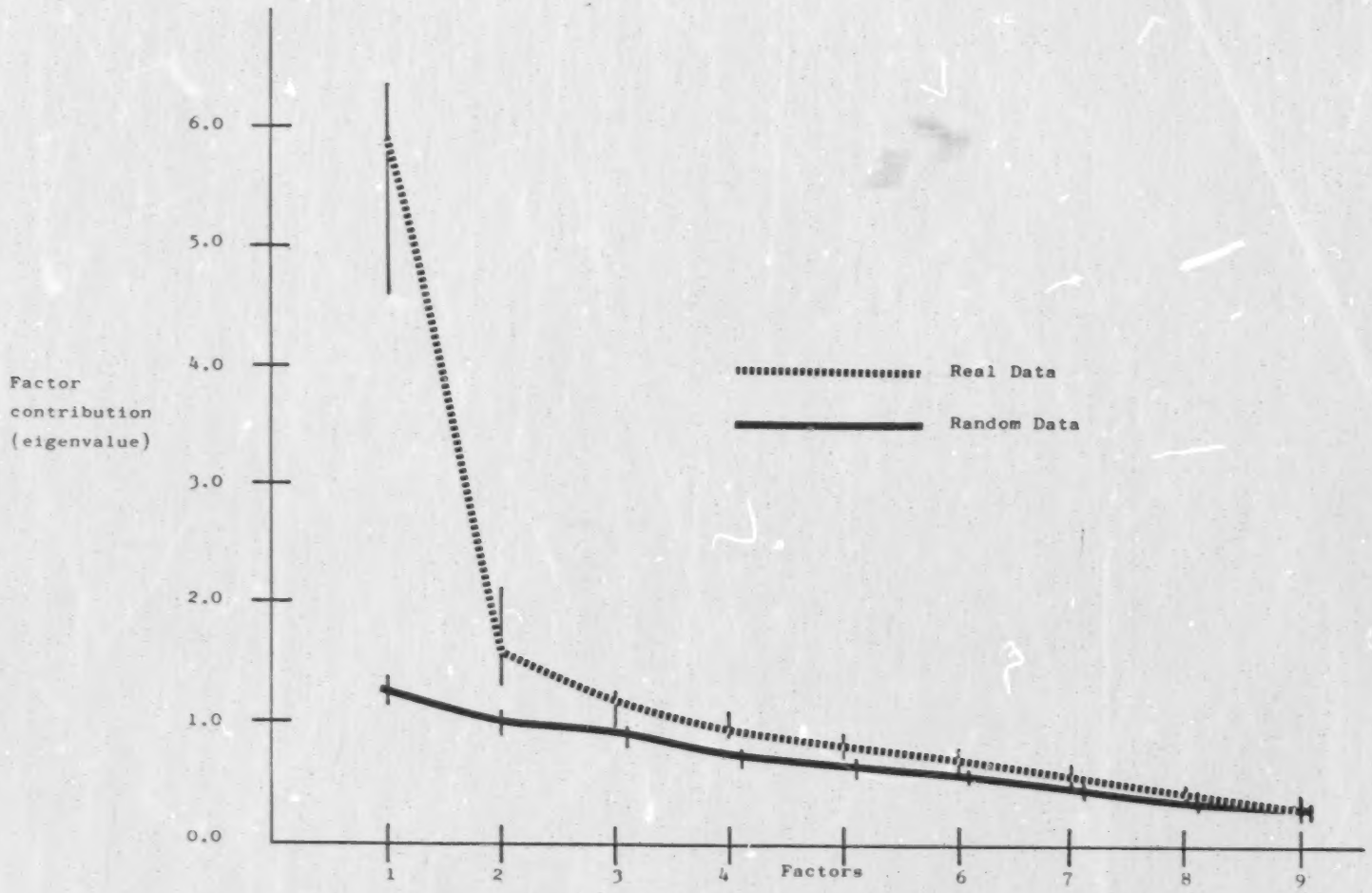

Figure 1. Factor contributions (eigenvalues) of the first nine principal factors extracted from each of six real data matrices and six random data matrices. For each factor, the vertical line spans the range of eigenvalues obtained. Horizontal lines connect the medians of each successive factor.

Factor structure. Table 14 lists the loadings (sorted in increasing order) of each subtest item on the first principal factor of each of the six real data subtests. Without exception, each of the 120 items sampled to comprise the six subtests loaded in the same direction on the first factor in its subtest, and each of their 120 factor loadings was statistically significant (Harman, 1960). These data meet the usual criterion for a "general" factor. That this factor structure is not found in data arising from more than one common factor is demonstrated by the factor loadings found in the six random data "subtests", also shown in Table 14. In the random data the first common factor was bipolar; i.e., significant positive and negative loadings occurred as consistently on the first factor as on lesser ones. In the real data, bipolarity was observed only in the second and lesser factors. Thus, the predictions made on the basis of the unidimensionality assumption were borne out in the real data, with no similar findings in the random data.

Item discriminations and first factor loadings. Table 15 lists, for each subtest, the factor loadings for the items in the subtest, and the correlation of each item with total score on the norming test (item discrimnation). Also shown are the product-moment correlations between each item's first factor loading and its discrimination index in the norming test. Correlations between factor loadings and item discrimination ranged from a low of .76 for test 1 to a high of .94 for test 5; the median correlation was .87. For all 120 items in the six subtests combined, the correlation between item discriminations and first factor loadings was .85. These results indicate a high degree of correspondence between the two sets of item characteristics. These data fulfill prediction 4, above, and are supportive of the hypothesis of a unidimensional set of items comprising the final item pool.

Summary. The results reported above are all characteristic of a set of unidimensional data; the same results would be extremely unlikely in cases in which a multidimensional latent space underlies individuals' responses to the test items. Thus, multidimensional data would be unlikely to produce results having the characteristics of the real data analyzed above: 1) highly similar factor structures in all six real subtests, characterized by a predominant first factor and subsequent factors similar to those of random data; 2) uniformly unidirectional and significant first-factor loadings of all 120 test items; and 3) high correlation between those factor loadings and the independently estimated item discriminations. These data, considered in combination, provide reason to believe that the latent

Table 14

Sorted factor loadings (f) of 20 test items on the first common factor extracted from each of the six real data correlation matrices, and factor loadings for the first factors of the six random data matrices

Real Data[a]												Random Data						
Subtest												"Subtest"						
1		2		3		4		5		6		1		2	3	4	5	6
item	f	item	f	item	f	item	f	item	f	item	f	item	f	f	f	f	f	f
58	.28	107	.26	148	.23	234	.35	151	.29	176	.26	1	.44	.50	.55	.47	.38	.48
159	.31	193	.32	93	.32	26	.35	217	.42	137	.32	2	.40	.40	.32	.35	.35	.46
168	.35	152	.35	53	.36	238	.37	133	.42	141	.36	3	.35	.25	.30	.23	.30	.44
157	.36	237	.38	81	.38	164	.37	111	.43	205	.37	4	.32	.20	.17	.23	.23	.42
131	.36	231	.40	214	.41	147	.37	224	.47	134	.41	5	.31	.10	.17	.21	.23	.19
233	.36	185	.44	132	.43	82	.41	110	.47	222	.41	6	.21	.10	.16	.17	.19	.17
94	.46	162	.45	39	.44	183	.41	208	.48	216	.42	7	.19	.04	.14	.16	.18	.12
202	.49	115	.50	207	.47	50	.47	95	.48	108	.42	8	.18	-.01	.11	.15	.06	.08
46	.51	140	.50	117	.49	114	.50	69	.49	133	.45	9	.15	-.02	.07	.14	.02	.08
211	.54	123	.52	145	.50	76	.54	112	.50	83	.45	10	.13	-.02	.00	.07	-.00	.03
203	.54	174	.53	235	.52	149	.54	209	.52	134	.46	11	.12	-.05	.00	.06	-.07	.02
59	.60	106	.53	63	.53	52	.55	154	.53	215	.50	12	.09	-.05	-.05	-.10	-.09	-.01
43	.61	169	.54	116	.53	143	.55	188	.56	204	.51	13	.06	-.06	-.12	-.10	-.17	-.03
47	.65	31	.57	80	.58	34	.56	60	.56	37	.52	14	.05	-.11	-.18	-.16	-.21	-.04
180	.66	56	.58	86	.65	227	.56	128	.63	232	.52	15	.05	-.12	-.22	-.22	-.24	-.05
134	.70	173	.59	85	.67	158	.57	103	.64	239	.57	16	.04	-.28	-.23	-.24	-.26	-.05
36	.71	109	.60	130	.70	104	.63	199	.65	113	.57	17	-.03	-.29	-.29	-.25	-.31	-.05
90	.73	84	.65	13	.74	221	.64	91	.67	156	.59	18	-.26	-.32	-.32	-.28	-.35	-.11
190	.76	66	.67	191	.76	44	.71	161	.68	125	.62	19	-.33	-.35	-.34	-.38	-.35	-.13
166	.81	127	.70	87	.77	101	.76	194	.74	27	.68	20	-.43	-.56	-.42	-.41	-.42	-.47

[a] Loadings for these factors have been reflected.

Table 15

Comparison of the first factor loadings (f) of the six real data 20-item subtests with their respective biserial correlation with total score (r) on the full-length norming test, and the correlation of f and r within each subtest

Subtest 1		2		3		4		5		6	
f	r	f	r	f	r	f	r	f	r	f	r
283	434	260	327	230	333	347	456	286	375	259	320
306	339	317	326	315	430	355	299	417	385	320	371
351	348	351	450	355	464	366	396	420	368	363	385
359	302	377	324	380	376	367	378	426	436	369	427
360	488	404	409	408	389	372	358	468	477	409	385
365	424	444	496	425	337	406	445	471	500	412	474
459	438	453	495	442	308	413	447	476	494	416	346
493	495	499	407	468	516	472	451	477	451	424	424
507	554	500	458	490	461	500	599	487	446	449	367
537	520	516	557	501	507	538	487	497	464	454	612
544	547	525	538	519	490	544	557	520	538	457	556
602	560	526	524	527	537	545	558	527	551	502	433
613	671	544	550	531	402	551	608	562	645	511	588
650	657	566	552	581	619	560	595	563	538	517	554
664	529	580	600	649	609	564	579	625	633	519	508
696	693	590	538	672	652	565	699	640	663	566	610
711	776	603	665	700	599	630	563	651	673	567	520
731	635	648	819	735	838	639	543	665	638	586	543
775	825	671	723	764	813	715	702	682	651	619	739
809	438	697	680	772	702	764	829	737	803	680	776
r=.76		r=.88		r=.87		r=.89		r=.94		r=.84	

Note.--Decimal points have been omitted.

space underlying the 120 items sampled is one-dimensional. By inference to the population of 369 items from which the 120 items were sampled, it is reasonable to treat the entire final item pool as unidimensional.

LIMITATIONS OF THE FINAL ITEM POOL

The research reported above resulted in the selection of an essentially homogeneous pool of 369 word knowledge test items, each characterized by estimated values of its classical and normal ogive model difficulty and discrimination parameters. While this item pool has some utility for adaptive testing research (e.g., Betz & Weiss, 1973; Weiss, 1973; Larkin & Weiss, 1974) it has a number of technical limitations. These limitations can be summarized in four categories: 1) generalizability; 2) traditional item parameters; 3) normal ogive parameters; and 4) factor analyses.

Generalizability

The use of a college population as the source of experimental subjects raises the problem of restriction of range. This is especially true when "ability" is being measured. All the norming done within the research reported here used college undergraduates as subjects; therefore, the results reported--particularly the item parameters--are obviously not generalizable beyond the population sampled. The difficulty and discrimination indices of the items, and perhaps even the factor structure, might vary considerably in less homogenous groups. The purpose of the norming work, however, was to establish a pool of items of known parameters from which to construct tests for adaptive administration by computer. The parent research project will use the obtained item parameters in research tests whose subjects will be drawn from the same sources sampled from here. Any other applications of the items normed herein will have to be preceded by separate norming studies on the population in question.

Availability of the normal ogive item parameters might serve to increase the generality of the item pool. At hand is a large set of test items calibrated on a sample of college students. Calibration of the normal ogive parameters of the items by the methods described above involved arbitrarily determining the metric for ability (θ) so that its mean was equal to 0.0 and its variance equal to 1.0. To calibrate the same set of items on another group, both the original norm group (group 1) and the new group (group 2) can be considered to be subgroups from a larger population; this assumption permits taking advantage of the invariance properties of the normal ogive item parameters.

Using the assumptions of normal ogive item parameter theory it is possible to calibrate the entire item pool for the new group by norming only a subset of the final item pool on that group. The parameters of the items in the subset would be estimated from the norming data obtained from group 2, just as was done for the original norm group (group 1). The newly-obtained item parameters of the subset items could be plotted against their parameters in group 1. Plotting the discrimination parameters should result in a linear relationship with a zero intercept. Plotting the difficulty parameters the same way should also yield a linear relationship, but not necessarily a zero intercept (Lord & Novick, 1968, pp. 380-381). From these data the two linear regression equations for the item parameters of group 1 on group 2 may be calculated. Application of these procedures to a subset of the items in the current pool would result in regression equations for transforming the present item pool parameters into parameters appropriate to a new group of subjects. The regression equation obtained in a subset of the item pool may be used as the basis for transformations of the parameters of every item in the pool. Or, if the new group were norming a new set of items for the first time, the newly obtained item parameters may be transformed into parameters comparable with those of the final item pool and relevant to group 1, the original "standard" group.

Traditional Item Parameters

Calculation of the traditional item discrimination indices was done using a different total score in each of the four norming tests. Since the content of the four tests was independent of the content of each of the others (with the exception of the calibration items), it can be argued that the discrimination indices are not comparable from one test to another. The possible lack of comparability of the biserial-r discrimination indices obtained from total scores on different, non-parallel, tests is an important consideration in generalizing the results of the analyses reported in this paper. Lacking any common reliable criterion across test administration, however, there was little choice but to assume comparability. This assumption is tenable, however, since a unitary trait was considered to underly the reliable variance in separate tests, and an individual's test score was expected to be an increasing function of his level on that trait. The results of the factor analyses above lend direct support to the assumption of a unitary trait. Futhermore, the same results lend indirect support to the comparability of item-test correlations across norming tests, in light of the high correspondence between each item's first factor loading and its correlation with score on the long norming test.

Normal Ogive Item Parameters

Stable estimates of the normal ogive item characteristic curve parameters require large numbers of examinees at each of several adjacent levels of θ. As a consequence, very large sample sizes are required for satisfactory estimates. The sample sizes reported herein are too small by usually specified criteria to yield anything but approximate values. Furthermore, the method used here employs the item-test biserial correlation as an approximation to the item-trait biserial correlation to estimate item parameters "a" and "b".

In the case of a purely unitary trait, the loading of an item variable on the single common factor extracted from the matrix of tetrachoric inter-item correlations is nearly identical with the item-trait biserial correlation (Henrysson, 1962). In the case of empirical data, that correspondence will be attenuated somewhat, but the item loadings on the first common factor (in data in which one common factor accounts for most of the variance) are satisfactory estimates of the item-trait biserial correlations (Bock & Wood, 1971). Inspection of the factor loading tables reported above reveals that, for those items selected randomly for factor analysis, the item-test biserial estimates corresponded closely to the first factor loadings. This correspondence was taken as support for the hypothesis of a unitary trait. These data can also be taken as justification for the approximation used in computation of the parameters of the item characteristic curves.

While it is recognized that the normal ogive item parameters reported here are approximations with an unknown degree of accuracy, these approximations appear to have characteristics of more exact latent trait parameter estimates, and are of practical utility in scaling items for adaptive test applications (e.g., Betz & Weiss, 1973; Larkin & Weiss, 1974).

Factor Analyses

The results of the separate factor analyses of six 20-item subsets have been taken as supporting evidence for the essential unidimensionality of the final item pool. One obvious limitation is that a larger subset--perhaps even the entire pool--should have been used. The considerations of sample size that led to rejection of such an approach have been discussed previously. It was felt that six replications, using random sampling procedures to select the 20-item subsets, should suffice to reveal any chance contributions to the factor structure due to item-sampling. At the same time, the use of multiple replications attributes a degree of generality to the

results which is not possible when the factor analysis is based on only one large set of items.

A DESIGN FOR FUTURE NORMING

Each of the technical limitations of the norming efforts reported above arose out of practical considerations imposed in part by the small sample sizes available in the college setting. Although the item parameters obtained by the present research are expected to be of practical utility in evaluating different strategies of adaptive testing, the present experience has provided a basis for the design of a more comprehensive norming study in which the shortcomings of the initial norming work may be overcome. Since a new norming study would necessarily precede the employment of any adaptive testing strategy in the evaluation of individual differences in any practical application, a plan for such a study is outlined below, along with a number of special considerations peculiar to the problems of developing a unidimensional item pool for computer-administered adaptive testing.

One such consideration is the size of the final item pool: it must be large enough to permit a number of items to be drawn from any one difficulty level without exhausting the items available at that level. The exact number of items will vary as a function of the planned length of the adaptive test and the heterogeneity of ability of the population of interest. Suffice it to say that the 369-item pool developed above is not too large. The relatively large size of the final item pool requires that the development item pool be even larger--perhaps twice as many items will be tried than will be satisfactory. Since it would be unreasonable to require a norming sample to take, say, a 600-item norming test, the development item pool will have to be divided into several different short norming tests, to be administered separately to different groups of testees.

In order to avoid problems in comparing these different groups, some common nucleus of items should be administered to each, such as calibration subtest II, which was developed for such a purpose. Another problem arising from giving different tests to different groups is that of the criterion against which the index of item discrimination will be calculated. It is highly desirable to have a common criterion for that purpose. The calibration subtest can also be useful for that purpose. As presently conceived, however, that subtest is too short to exhibit the reliability desirable in such a criterion. Therefore, a longer criterion subtest is proposed, which would serve as a standard for intergroup ability comparisons, as a

criterion for computing discrimination indices of norming-test items, and as a basis for scaling transformations of item parameters (see below). Had such a subtest been available at the outset of the present research it could have mitigated several of the technical problems which were inherent in the present work.

Other limitations of the present work were directly due to the relatively small number of subjects employed in the norming. The proposal below is predicated on the use of considerably larger norming groups. Large samples will permit direct factor analysis of the test data, and hence a direct evaluation of the dimensionality of the item pool. Further, it will permit the use of "hold-out" groups for cross-validation or replication of results in order to distinguish stable results from sampling fluctuations. Finally, large sample sizes will permit stable estimates of latent-trait model item parameters, which promise several advantages for adaptive testing.

Listed below in outline form are the basic steps of the proposal. Each of the seven steps is discussed in some detail following the outline itself.

Basic steps

1. Define the population of interest.
2. Compile an initial development test and norm it on a representative sample from the population.
3. On the basis of the norming data, construct a calibration/criterion subtest.
4. Construct several long secondary development tests incorporating the calibration/criterion subtest.
5. Norm these development tests on large representative samples from the population.
6. Perform item analyses of the development tests, employing calibration/criterion subtest score as the criterion.
7. Select items for the item pool on the basis of the item analysis data.

Operationalization

Population definition. The definition of the population of interest will delimit the examinee groups in which

the item pool is to be employed. One of the premises favoring use of adaptive tests is that such tests can provide more information than conventional tests about examinees in the extremes of the distribution (Weiss & Betz, 1973); from this it follows that an adaptive test can be employed within a highly heterogeneous population. Whereas a conventional test is appropriately used within a rather narrowly defined population (e.g., high school seniors) a well-designed and constructed adaptive test should be useful in a much more broadly defined group (such as ninth to twelfth year high school students). The adaptive procedure itself, and the assumed invariance properties of the latent trait model item parameters, permit defining the population in terms of two or more subgroups whose average "ability" is known to differ. Taking advantage of that fact makes it possible--even desirable--to define the population of interest rather broadly, and once the final item pool is developed, to apply adaptive tests across a relatively broad range of subgroups.

Initial norming. The initial development test will be used solely to scale items for the purpose of selecting items for the calibration/criterion subtest. For that purpose the development items should span the range of item difficulty required for the calibration subtest (see below). This will require some judgment on the part of the test constructor, of course. The size of the initial development test should be two to three times the size of the calibration test which is to be selected from it, to provide some assurance of its yielding a sufficient number of items having the characteristics desired for the calibration subtest items.

Once the initial development test is assembled, it should be administered to a representative sample of persons from the general population of interest, or to samples from two or more of the subgroups which are subsumed under the general population. Sampling from the entire population is preferable, since the scaled item parameters so obtained would be relative to the entire population. This would facilitate constructing the calibration subtest to have the characteristics described below. However, if the second alternative is followed, the subjects should be selected to be representative of both average and extreme levels of ability in the population.

Development of the calibration test. On the basis of the item analysis data from the initial development test, the items for the calibration/criterion subtest should be selected. The subtest so constructed will serve as a "standard" in all subsequent norming, so great care is necessary in its construction. It should have high internal

consistency, and its items should span the entire range of item difficulty.

The subtest needs to be of at least moderate length in order to make it reliable. Since it will serve later as a standard criterion against which item discrimination will be calculated, those discrimination indices will be attenuated artifactually to the extent that the subtest is unreliable. With its psychometric characteristics held constant, a test with item intercorrelations in the typical range will rapidly approach its maximum reliability at a test length from 40 to 60 items. This length is therefore suggested for the calibration/criterion subtest.

A wide range of item difficulty is required in order to employ this subtest as a basis for transformations of item parameters from those obtained within some standard group to parameters appropriate to some other population subgroup. The parameters transformed may be the proportions correct (using the procedure suggested by Gulliksen 1950, pp. 377) or may be the normal ogive item parameters (according to the Lord and Novick (1968) method). Achieving stability of any such transformation is facilitated by avoiding restriction of range; in this case, restriction of range is avoided by exploiting the entire useful range of item difficulty.

Constructing the test from items spanning the useful range of difficulty would have the effect of attenuating the internal consistency reliability of the subtest. Item difficulties around .50 would tend to maximize reliability (Nunnally, 1967), and it is thus desireable to incorporate into the test a number of items of median difficulty. High internal consistency as well as the desired range of difficulty could be achieved by combining two subsets of items to comprise the test. One subset of items, about half the number in the subtest, should have a rectangular distribution of item difficulties. A second subset of items should be "peaked" at median difficulty

These considerations imply some specific suggestions for constructing the calibration/criterion subtest on the basis of item analysis data.

1) Establish subtest length, K, between 40 and 60 items to permit reliability to approach its maximum.

2) Consider the items in the range of p-values from .05 to .95.

3) Divide the p-value range into ten equal intervals and select the K/20 highest discriminating items from within each interval. This will yield K/2 highly discriminating items with an approximately rectangular difficulty distribution.

4) Select the remaining K/2 items from the most discriminating items of median difficulty value. This will yield an approximately peaked and highly discriminating subset of items.

5) Using a holdout sample from the norming data, analyze the K items as a subtest. Make refinements as necessary to provide high internal consistency and the desired distribution of item difficulty. This analysis should include an investigation of the dimensionality of the subtest. The calibration/criterion subtest should be considered complete only after it has been shown to measure a unidimensional trait as well as to have the desired psychometric characteristics.

Development of a trial item pool. After the development of the calibration/criterion subtest, calibration of a large item pool may proceed. Since a large final item pool is required, several long norming tests should be constructed, each of which will be administered to norming groups in conjunction with the calibration subtest. An attempt should be made to make these norming tests approximately parallel in terms of overall difficulty and order of item difficulty within each test. In other words, the tasks of the examinees should be comparable from one norming test to another, within the ability of the test constructors to make them so. The items of the calibration subtest may be incorporated as part of the norming tests or administered separately at the same sitting.

Norming of the trial item pool. Norming of the development test items must be accomplished with very large samples from the population, in order to achieve stability of the item characteristic curve parameters. These parameters are typically estimated by a procedure which divides the criterion score range into fractiles, and utilizes the obtained proportion correct within each fractile. The number of individual scores within each fractile has important implications for the power of the statistical estimation of the item characteristic curve parameters.

To permit the analysis of the unidimensionality of the trial item pool, and the resulting final item pool, items must be administered in sufficient combinations so that it is possible to compute a total item intercorrelation matrix. Thus, if item sampling is used to obtain large numbers of subjects on smaller numbers of items, the sampling must be designed with sufficient redundancy so that the intercorrelations of all items in the pool are determinable on samples of roughly equivalent sizes.

Item analyses. Item analysis of the norming tests will employ the examinee's score on the calibration/criterion subtest as the criterion against which item discrimination indices are computed. As discussed earlier, this has two advantages. First, it makes every item-criterion correlation comparable with every other one, regardless of the subsets in which the items were normed. Second, by providing a comparable ability estimate for every testee, it legitimizes the use of the item-subtest biserial correlation as an estimate of the item-ability biserial correlation which is the basis for calculating the normal-ogive item parameters.

Analyses of the kind recommended here can be readily accomplished with existing computer programs, such as those given by Kolakowski and Bock (1972) and by Baker (1969). Baker's program permits analysis of any subset of items, with total score on that subtest used as an external criterion score for a second analysis of all the items.

Selection of the final item pool. On the basis of the item analysis data from the norming tests, selection of a final item pool for adaptive testing may be accomplished. Highly discriminating items are preferable, whose difficulty parameters span the range of abilities likely to occur within any subgroup of the population. For this purpose the item difficulty (b) range of -3.00 to +3.00 should be well represented, with items distributed approximately equally across the range. Items with very low discrimination parameters ("a" less than .20) will be of little value for most measurement purposes.

References

Alf, E. & Abrahams N. A significance test for biserial r. Educational and Psychological Measurement, 1971, 31, 637-640.

American Council on Education. Psychological Examination for College Freshmen. Educational Testing Service, 1947.

American Council on Education. Scholastic Aptitude Test for Students in the 11th Grade. Educational Testing Service, 1952.

Baker, F. B. FORTAP: a FORTRAN test analysis package. Laboratory of experimental design, Department of Educational Psychology, University of Wisconsin, 1969.

Betz, N. E. & Weiss, D. J. An empirical study of computer-administered two-stage ability testing. Research Report 73-4, Psychometric Methods Program, Department of Psychology, University of Minnesota, 1973.

Bock, R. D. & Wood, R. Test theory. Annual Review of Psychology, 1971, 22, 193-219.

Carroll, J. B. The nature of the data, or how to choose a correlation coefficient. Psychometrika, 1961, 26, 347-372.

Cooperative Test Service. Cooperative English Test, form S. American Council on Education, 1942.

Cooperative Test Service. Cooperative English Test, form T. American Council on Education, 1943.

DeWitt, L. J. & Weiss, D. J. A computer software system for adaptive ability measurement. Research Report 74-1, Psychometric Methods Program, Department of Psychology, University of Minnesota, 1974.

Gulliksen, H. Theory of Mental Tests. New York: Wiley, 1950.

Harman, H. H. Modern factor analysis. Chicago: University of Chicago Press, 1960.

Henrysson, S. The relationship between factor loadings and biserial correlations in item analysis. Psychometrika, 1962, 27, 419-424.

Indow, T. & Samejima, F. On the results obtained by the absolute scaling model and the Lord model in the field of intelligence. Third report, the Psychology Laboratory on the Hiyoshi Campus, Keio University, Japan, 1966. (In English)

Kolakowski, D. & Bock, R. D. LOGOG: a FORTRAN IV program for maximum likelihood item analysis and test scoring: logistic model for multiple item responses. Research Memorandum Number 13, Statistical Laboratory, Department of Education, University of Chicago, 1972.

Larkin, K. C. & Weiss, D. J. An empirical investigation of computer-administered pyramidal ability testing. Research Report 74-3, Psychometric Methods Program, Department of Psychology, University of Minnesota, 1974.

Lord, F. M. A theory of test scores. Psychometric Monograph, 1952, No. 7.

Lord, F. M. & Novick, M. R. Statistical theories of mental test scores. Reading, Massachusetts: Addison-Wesley, 1968.

McBride, J. R. TETREST: a FORTRAN IV program for calculating tetrachoric interitem correlations from scored item response data. Unpublished program. Department of Psychology, University of Minnesota, 1974.

Nunnally, J. C. Psychometric Theory. New York: McGraw-Hill, 1967.

Owen, R. J. A Bayesian approach to tailored testing. Research Bulletin 69-92, Educational Testing Service, Princeton, N.J., 1969.

Tucker, L. R. Maximum validity of a test with equivalent items. Psychometrika, 1946, 11, 1-13.

Urry, V. A monte carlo investigation of logistic mental test models. Unpublished doctoral dissertation, Purdue University, 1970. (University Microfilms, No. 71-9475, Ann Arbor, Mich.)

Weiss, D. J. & Betz, N. E. Ability measurement: conventional or adaptive? Research Report 73-1, Psychometric Methods Program, Department of Psychology, University of Minnesota, 1973.

Weiss, D. J. The stratified adaptive computerized ability test. Research Report 73-3, Psychometric Methods Program, Department of Psychology, University of Minnesota, 1973.

Weiss, D. J. Strategies of adaptive ability measurement. Research Report 74-x, Psychometric Methods Program, Department of Psychology, University of Minnesota, 1974 (in preparation).

Appendix A

Traditional and normal ogive parameters for the items of the norming item pool

ITEM	N	POINT BISERIAL	BISERIAL R	PROPORTION CORRECT	DIFFICULTY (B)	DISCRIMINATION(A)
1	192	.033	.122	.989	-18.774	.123
2	192	.180	.441	.960	-4.200	.491
3	192	.013	.042	.984	-51.057	.042
4	192	.114	.359	.984	-5.973	.385
5	192	.308	.565	.921	-2.499	.685
7	192	.310	1.134	.989	-2.414	3.000
8	192	.185	.680	.990	-3.421	.927
9	192	.301	.790	.974	-2.460	1.289
10	192	.060	.218	.989	-10.506	.223
11	192	.226	.829	.990	-2.806	1.482
12	192	.081	.298	.990	-7.807	.312
13	192	.437	.838	.932	-1.779	1.536
14	192	.237	.873	.990	-2.665	1.790
15	192	.104	.209	.942	-7.521	.214
16	192	.232	.573	.969	-3.257	.599
17	192	.239	.560	.963	-3.190	.676
18	192	.159	.416	.974	-4.671	.457
19	192	.156	.571	.989	-4.011	.696
20	192	.110	.348	.984	-6.162	.371
21	192	-.024	-.113	.995	22.795	-.114
22	192	.351	.731	.948	-2.224	1.071
23	192	.151	.553	.990	-4.207	.664
24	192	.268	.846	.984	-2.535	1.587
25	192	.224	1.074	.995	-2.715	3.000
26	192	.226	.299	.709	-1.841	.313
27	192	.405	.776	.932	-1.921	1.230
28	192	.222	1.067	.995	-2.715	3.000
29	192	.125	.292	.964	-6.161	.305
30	192	.169	.278	.885	-4.316	.289
31	192	.307	.552	.916	-2.498	.662
32	192	.170	.538	.984	-3.986	.638
33	192	.414	.540	.677	-.851	.642
34	192	.370	.595	.875	-1.933	.740
35	192	.178	.224	.568	-.765	.230
36	192	.543	.776	.800	-1.085	1.230
37	192	.431	.554	.649	-.691	.665
38	192	.048	.229	.995	-11.248	.235
39	192	.177	.308	.906	-4.274	.324
40	192	.480	.714	.831	-1.342	1.020
41	192	.107	.252	.964	-7.139	.260
42	192	.222	.067	.995	-2.715	3.000
43	192	.475	.671	.791	-1.207	.905
44	192	.428	.702	.885	-1.710	.986
45	192	.110	.169	.853	-6.209	.171
46	192	.426	.554	.674	-.814	.665
47	192	.457	.657	.805	-1.308	.871
48	192	.163	.234	.305	-3.674	.241
49	192	.140	.177	.603	-1.475	.180
50	192	.359	.451	.542	-.234	.505
51	192	.409	.756	.843	-1.332	1.155
52	192	.411	.518	.558	-.282	.606
53	192	.356	.404	.681	-1.014	.524
54	192	.228	.288	.577	-.674	.301
55	192	.137	.264	.932	-5.647	.274
56	192	.476	.600	.568	-.285	.750

ITEM	N	POINT BISERIAL	BISERIAL R	PROPORTION CORRECT	DIFFICULTY (B)	DISCRIMINATION(A)
57	192	.158	.227	.194	3.803	.233
58	192	.335	.434	.661	-.957	.482
59	328	.428	.537	.463	.173	.637
60	192	.440	.553	.447	.241	.664
61	192	-.003	-.008	.974	242.892	-.008
62	192	.133	.378	.979	-5.380	.408
63	192	.303	.537	.911	-2.508	.637
64	192	.260	.954	.990	-2.452	3.000
65	192	.243	.691	.979	-2.943	.956
66	192	.333	.623	.926	-2.322	.796
67	192	.114	.219	.932	-6.808	.224
68	192	.276	.682	.969	-2.737	.933
69	192	.214	.447	.948	-3.637	.500
70	192	.306	.756	.969	-2.469	1.155
71	192	.258	.944	.989	-2.414	3.000
72	192	.113	.252	.958	-6.857	.260
73	192	.224	.367	.884	-3.257	.395
74	192	.198	.343	.906	-3.838	.365
75	192	.009	.032	.990	-72.698	.032
76	192	.314	.487	.857	-2.191	.558
77	192	.191	.385	.942	-4.083	.417
78	192	.136	.388	.979	-5.241	.421
79	192	.119	.215	.917	-6.443	.220
80	192	.306	.619	.943	-2.553	.788
81	192	.238	.376	.866	-2.946	.406
82	192	.267	.445	.891	-2.768	.497
83	192	.389	.612	.864	-1.795	.774
84	192	.417	.819	.937	-1.868	1.427
85	328	.448	.605	.741	-1.068	.760
86	192	.411	.609	.828	-1.554	.768
87	192	.502	.702	.779	-1.095	.986
88	328	.359	.531	.823	-1.745	.627
89	192	.276	.558	.942	-2.817	.672
90	192	.412	.635	.853	-1.653	.822
91	192	.495	.637	.647	-.592	.826
92	192	-.010	-.022	.947	73.474	-.022
93	192	.268	.430	.875	-2.675	.476
94	192	.321	.438	.754	-1.569	.487
95	192	.300	.451	.839	-2.196	.505
96	328	.386	.751	.921	-1.880	1.137
97	192	-.009	-.013	.172	-72.792	-.013
98	192	.064	.081	.508	-.248	.081
99	328	.235	.782	.985	-2.775	1.255
100	192	.218	.487	.958	-3.548	.558
101	328	.432	.715	.884	-1.672	1.023
102	192	.257	.943	.990	-2.452	3.000
103	192	.457	.663	.813	-1.341	.886
104	192	.445	.563	.589	-.400	.681
105	192	.256	.674	.974	-2.883	.912
106	192	.310	.524	.895	-2.392	.615
107	192	.253	.327	.346	1.211	.346
108	192	.307	.424	.766	-1.712	.468
109	192	.484	.665	.759	-1.057	.890
110	192	.380	.500	.698	-1.037	.577
111	192	.344	.435	.421	.458	.483

ITEM	N	POINT BISERIAL	BISERIAL R	PROPORTION CORRECT	DIFFICULTY (B)	DISCRIMINATION(A)
112	192	.345	.462	.726	-1.300	.521
113	328	.413	.520	.449	.247	.609
114	328	.474	.612	.345	.652	.774
115	192	.291	.407	.222	1.881	.446
116	328	.281	.354	.553	-.376	.379
117	192	.348	.461	.708	-1.188	.519
118	192	.075	.106	.209	7.641	.107
119	192	.224	.286	.366	1.197	.298
120	192	.433	.583	.266	1.072	.718
121	192	.245	.574	.963	-3.113	.701
122	192	.151	.724	.995	-3.558	1.050
123	192	.420	.557	.711	-.999	.671
124	192	.250	.710	.979	-2.864	1.008
125	192	.367	.739	.942	-2.127	1.097
126	192	.306	.660	.953	-2.537	.879
127	192	.428	.680	.870	-1.656	.927
128	192	.485	.634	.682	-.747	.820
129	192	.447	.734	.885	-1.635	1.081
130	192	.456	.599	.695	-.852	.748
131	192	.261	.488	.927	-2.979	.559
132	192	.197	.337	.900	-3.803	.358
133	192	.303	.379	.513	-.086	.410
134	192	.353	.693	.937	-2.208	.961
135	192	.220	.367	.890	-3.342	.395
136	192	.200	.258	.654	-1.535	.267
137	192	.292	.371	.608	-.739	.400
138	192	.339	.836	.968	-2.216	1.524
139	192	.307	.385	.471	.189	.417
140	192	.343	.458	.276	1.299	.515
141	192	.281	.385	.759	-1.826	.417
142	192	.198	.255	.651	-1.522	.264
143	192	.474	.608	.635	-.568	.766
144	192	.423	.531	.539	-.184	.627
145	192	.401	.507	.583	-.413	.588
146	192	.414	.519	.500	.000	.607
147	192	.277	.358	.340	1.152	.383
148	192	.265	.333	.445	.415	.353
149	192	.424	.557	.693	-.906	.671
150	192	.093	.137	.178	6.737	.138
151	192	.228	.376	.885	-3.192	.406
152	327	.353	.481	.251	1.396	.549
153	192	.008	.011	.293	49.513	.011
154	192	.432	.552	.626	-.582	.662
155	192	.247	.320	.667	-1.349	.338
156	192	.424	.543	.634	-.631	.647
157	192	.236	.302	.628	-1.081	.317
158	192	.345	.699	.943	-2.261	.977
159	192	.262	.339	.337	1.241	.360
160	192	.098	.133	.257	4.907	.134
161	192	.519	.653	.566	-.255	.862
162	327	.348	.460	.295	1.171	.518
163	192	-.006	-.010	.135	-110.306	-.010
164	192	.299	.378	.408	.616	.408
165	192	.222	.278	.537	-.334	.289
166	327	.340	.538	.137	2.033	.638

ITEM	N	POINT BISERIAL	BISERIAL R	PROPORTION CORRECT	DIFFICULTY (B)	DISCRIMINATION(A)
167	192	.181	.233	.363	1.504	.240
168	192	.267	.348	.318	1.360	.371
169	192	.031	.040	.385	7.309	.040
170	192	.121	.174	.195	4.940	.177
171	192	.050	.070	.230	10.555	.070
172	192	.201	.252	.453	.469	.260
173	326	.418	.605	.807	-1.433	.760
174	192	.400	.538	.267	1.156	.638
175	192	.100	.159	.132	7.025	.161
176	192	.251	.320	.613	-.897	.338
177	192	.139	.180	.330	2.444	.183
178	192	.113	.169	.164	5.788	.171
179	192	.168	.211	.547	-.560	.216
180	326	.279	.398	.205	2.070	.434
181	192	.261	.686	.974	-2.833	.943
182	192	.155	.567	.989	-4.039	.688
183	326	.390	.512	.684	-.935	.596
184	192	.303	.556	.921	-2.539	.669
185	192	.372	.496	.720	-1.175	.571
186	192	.428	.678	.868	-1.647	.922
187	192	.219	.410	.926	-3.528	.450
188	326	.455	.579	.608	-.473	.710
189	192	.380	.550	.811	-1.603	.659
190	192	.458	.825	.917	-1.679	1.460
191	192	.488	.813	.890	-1.509	1.396
192	192	.106	.248	.963	-7.204	.256
193	192	.233	.326	.220	2.369	.345
194	192	.534	.803	.839	-1.233	1.347
195	192	.096	.121	.422	1.626	.122
196	192	.387	.869	.958	-1.988	1.756
197	192	.161	.217	.740	-2.965	.222
198	192	.277	.596	.953	-2.810	.742
199	192	.455	.677	.832	-1.421	.920
200	192	.069	.094	.758	-7.446	.094
201	192	.179	.273	.848	-3.765	.284
202	192	.290	.495	.899	-2.578	.570
203	192	.420	.547	.677	-.840	.653
204	192	.431	.588	.750	-1.147	.727
205	192	.337	.427	.604	-.618	.472
206	192	.317	.710	.958	-2.434	1.008
207	192	.406	.516	.607	-.526	.602
208	192	.393	.503	.634	-.681	.582
209	192	.425	.537	.586	-.405	.637
210	192	.091	.118	.335	3.611	.119
211	192	.404	.520	.646	-.720	.609
212	192	.188	.465	.969	-4.014	.525
213	192	.217	.274	.429	.653	.285
214	192	.275	.389	.791	-2.082	.422
215	192	.340	.433	.389	.651	.480
216	192	.271	.346	.375	.921	.369
217	192	.300	.393	.312	1.247	.427
218	192	.247	.315	.315	-.928	.332
219	192	.091	.155	.101	8.231	.157
220	192	-.172	-.370	.047	-4.526	-.398
221	192	.420	.543	.656	-.740	.647

ITEM	N	POINT BISERIAL	BISERIAL R	PROPORTION CORRECT	DIFFICULTY (B)	DISCRIMINATION(A)
222	192	.362	.474	.686	-1.022	.538
223	192	.044	.062	.226	12.130	.062
224	192	.371	.477	.646	-.785	.543
225	192	.126	.165	.316	2.903	.167
226	192	.052	.067	.363	5.231	.067
227	192	.391	.579	.827	-1.628	.710
228	192	.055	.078	.209	10.383	.078
229	192	-.072	-.104	.188	-8.512	-.105
230	192	.126	.163	.344	2.464	.165
231	192	.320	.409	.374	.786	.448
232	192	.353	.508	.806	-1.699	.590
233	192	.338	.424	.529	-.172	.468
234	192	.358	.456	.623	-.687	.512
235	192	.364	.490	.733	-1.269	.562
236	192	.124	.215	.095	6.096	.220
237	326	.200	.343	.098	3.770	.365
238	192	.312	.396	.398	.653	.431
239	192	.448	.610	.749	-1.101	.770
240	192	.206	.483	.964	-3.725	.552
241	83	.322	.441	.759	-1.594	.491
242	83	.233	.296	.386	.979	.310
243	83	.146	.187	.373	1.732	.190
245	166	.250	.355	.205	2.321	.380
246	83	-.079	-.142	.085	-9.663	-.143
247	83	.201	.264	.305	1.932	.274
248	83	.165	.207	.543	-.522	.212
249	83	.085	.124	.181	7.351	.125
250	83	.027	.035	.373	9.255	.035
252	83	.244	.307	.542	-.344	.323
253	192	.273	.364	.274	1.650	.391
254	83	.380	.508	.277	1.165	.590
255	83	.293	.506	.904	-2.578	.587
256	83	.257	.382	.831	-2.508	.413
257	83	.230	.288	.488	.104	.301
258	83	.300	.376	.500	.000	.406
259	83	.199	.250	.463	.372	.258
260	156	.228	.323	.317	1.474	.341
261	83	.195	.306	.928	-3.992	.393
262	83	.331	.571	.904	-2.285	.696
263	83	.338	.456	.265	1.377	.512
264	83	.518	.651	.446	.209	.858
265	83	.486	.611	.458	.173	.772
266	83	.521	.654	.458	.161	.865
267	83	.319	.400	.470	.188	.436
268	83	.157	.197	.494	.076	.201
269	83	.125	.186	.171	5.109	.189
270	83	.510	.653	.634	-.524	.862
271	83	.371	.466	.439	.329	.527
272	83	.559	.701	.537	-.132	.983
273	83	.311	.437	.217	1.790	.486
274	185	.272	.388	.204	2.133	.421
275	83	.084	.106	.590	-2.147	.107
276	150	.263	.377	.788	-2.121	.407
277	189	.201	.269	.270	2.278	.279
278	83	.115	.152	.293	3.583	.154

ITEM	N	POINT BISERIAL	BISERIAL R	PROPORTION CORRECT	DIFFICULTY (B)	DISCRIMINATION(A)
280	83	.047	.065	.232	11.266	.065
281	123	.194	.246	.403	.998	.254
282	83	.035	.063	.084	21.883	.063
283	83	.553	.695	.458	.152	.967
284	83	.007	.008	.422	24.597	.008
285	83	.409	.581	.795	-1.418	.714
286	83	.239	.306	.361	1.163	.321
287	83	.307	.403	.695	-1.266	.440
288	83	.371	.491	.293	1.109	.564
289	83	.264	.331	.506	-.045	.351
290	214	.224	.390	.094	3.376	.424
291	164	.305	.403	.299	1.308	.440
292	83	.339	.430	.598	-.577	.476
293	83	.368	.485	.699	-1.075	.555
294	83	.442	.575	.325	.789	.703
295	83	.342	.428	.506	-.035	.474
296	83	.533	.674	.410	.338	.912
297	208	.259	.371	.196	2.307	.400
298	214	.260	.398	.149	2.615	.434
299	83	.356	.464	.325	.978	.524
300	83	-.058	-.124	.048	-13.424	-.125
301	83	.481	.603	.482	.075	.756
302	83	.358	.451	.434	.369	.505
303	83	.204	.353	.096	3.696	.377
304	83	.299	.386	.349	1.005	.418
305	83	.077	.100	.659	-4.097	.101
306	161	.340	.440	.335	.969	.490
307	83	.379	.490	.659	-.836	.562
308	83	.189	.237	.469	.328	.244
309	83	.277	.429	.145	2.466	.475
310	83	.206	.262	.386	1.106	.271
311	83	.364	.551	.843	-1.827	.660
312	83	.230	.384	.108	3.222	.416
313	83	.435	.566	.675	-.802	.687
314	83	.075	.130	.096	10.036	.131
315	83	.507	.637	.458	.166	.826
316	83	.075	.105	.217	7.451	.106
317	83	.094	.125	.277	4.734	.126
318	83	.296	.373	.554	-.364	.402
319	83	.375	.526	.217	1.487	.618
320	83	.161	.204	.590	-1.115	.208
321	83	.410	.531	.337	.792	.627
322	83	.359	.504	.783	-1.552	.584
323	83	.146	.221	.157	4.556	.227
324	83	.278	.348	.488	.086	.371
325	83	-.037	-.054	.171	-17.597	-.054
327	83	.394	.496	.549	-.248	.571
328	83	.302	.475	.136	2.313	.540
329	83	.520	.654	.554	-.208	.865
330	83	.009	.013	.169	73.702	.013
331	83	-.103	-.139	.259	-4.651	-.140
332	83	.462	.604	.687	-.807	.758
333	83	.192	.242	.463	.384	.249
334	142	.064	.080	.303	6.447	.080
335	83	-.169	-.236	.222	-3.243	-.243

ITEM	N	POINT BISERIAL	BISERIAL P	PROPORTION CORRECT	DIFFICULTY (B)	DISCRIM-INATION(A)
336	83	.303	.442	.183	2.045	.493
337	83	.532	.700	.305	.729	.980
338	214	.167	.245	.178	3.767	.253
339	83	.050	.077	.146	13.685	.077
340	83	.486	.613	.427	.300	.776
341	125	.276	.349	.397	.748	.372
342	83	.488	.612	.458	.172	.774
343	83	.132	.168	.398	1.539	.170
344	83	.032	.053	.108	23.344	.053
345	83	.016	.021	.614	-13.798	.021
346	83	.021	.028	.265	22.429	.028
347	83	.581	.730	.458	.144	1.068
348	83	.041	.053	.325	8.562	.053
349	83	.448	.594	.711	-.937	.738
350	83	.174	.218	.458	.484	.223
352	83	.004	.005	.366	68.493	.005
353	83	.105	.149	.207	5.482	.151
354	83	.201	.254	.585	-.845	.263
355	83	.292	.369	.585	-.582	.397
356	83	.098	.123	.512	-.245	.124
357	83	.143	.183	.370	1.813	.186
358	83	.075	.099	.305	5.152	.099
359	83	.359	.503	.220	1.535	.582
360	187	.234	.321	.242	2.180	.339
361	83	.117	.150	.361	2.372	.152
362	83	.169	.214	.402	1.160	.219
363	83	-.037	-.050	.241	-14.062	-.050
364	202	.206	.309	.168	3.114	.325
365	83	.433	.552	.622	-.563	.662
366	83	.109	.175	.123	6.629	.178
367	115	.277	.353	.365	.978	.377
368	83	.227	.286	.420	.706	.298
369	83	.390	.490	.542	-.215	.562
370	83	-.011	-.015	.241	-46.873	-.015
371	147	.268	.358	.716	-1.595	.383
372	83	.206	.259	.542	-.407	.268
373	83	.083	.121	.185	7.409	.122
374	83	.206	.267	.333	1.617	.277
375	83	.346	.437	.420	.462	.486
376	209	.124	.184	.197	4.633	.187
377	83	.317	.398	.537	-.233	.434
378	160	.324	.438	.264	1.441	.487
379	83	.352	.542	.146	1.944	.645
380	83	.562	.708	.568	-.242	1.003
381	185	.320	.451	.210	1.788	.505
382	83	.424	.538	.602	-.481	.638
383	193	.254	.340	.268	1.820	.362
384	83	.056	.078	.217	10.030	.078
385	198	.265	.386	.182	2.352	.418
386	83	.456	.572	.469	.136	.697
387	83	.041	.054	.720	-10.793	.054
388	212	.246	.391	.132	2.857	.425
389	83	.149	.230	.145	4.601	.236
390	83	.412	.531	.651	-.731	.627
391	83	.342	.433	.590	-.526	.480

ITEM	N	POINT BISERIAL	BISERIAL R	PROPORTION CORRECT	DIFFICULTY (B)	DISCRIM-INATION(A)
392	83	.109	.140	.349	2.772	.141
393	83	.342	.443	.663	-.950	.494
394	83	.147	.187	.398	1.382	.190
395	83	.023	.034	.169	28.180	.034
396	83	.059	.074	.410	3.075	.074
397	83	.274	.349	.386	.830	.372
398	216	.316	.523	.111	2.335	.614
400	83	.246	.323	.301	1.615	.341
401	83	.140	.204	.819	-4.468	.208
402	83	.005	.009	.073	161.534	.009
403	121	.223	.286	.360	1.253	.298
501	67	.506	.680	.731	-.906	.927
502	68	.376	.474	.559	-.313	.538
503	66	.478	.610	.621	-.505	.770
504	63	.227	.294	.333	1.468	.308
505	71	.237	.299	.408	.778	.313
506	72	.377	.473	.486	.074	.537
507	63	.212	.340	.873	-3.355	.362
508	70	.015	.022	.171	43.192	.022
509	65	.055	.075	.246	9.162	.075
510	68	-.153	-.201	.309	-2.481	-.205
511	66	.157	.199	.394	1.351	.203
512	67	.188	.275	.179	3.342	.286
513	69	-.040	-.052	.328	-8.566	-.052
514	80	.291	.389	.275	1.535	.423
514	58	.227	.304	.275	1.966	.319
515	66	.479	.658	.759	-1.069	.874
516	72	.197	.247	.485	.152	.255
517	68	-.035	-.067	.069	-22.139	-.067
518	71	.010	.013	.353	29.018	.013
519	64	.292	.366	.535	-.240	.393
520	62	.196	.254	.328	1.754	.263
521	62	.244	.318	.323	1.444	.335
522	71	.482	.633	.690	-.783	.818
523	65	.400	.507	.400	.500	.588
524	68	.022	.033	.147	31.800	.033
525	68	.246	.314	.382	.956	.331
526	66	.389	.494	.394	.544	.568
527	63	.128	.315	.032	5.880	.332
528	69	.270	.385	.203	2.158	.417
529	67	.068	.088	.343	4.594	.088
530	74	-.189	-.307	.122	-3.795	-.323
531	68	.174	.220	.412	1.011	.226
532	67	-.300	-.658	.045	-2.577	-.874
533	70	.192	.253	.300	2.073	.262
534	67	.223	.303	.254	2.185	.318
535	67	.403	.523	.672	-.852	.614
536	65	-.019	-.029	.846	35.153	-.029
537	62	.199	.362	.081	3.863	.388
538	65	.451	.565	.477	.102	.685
540	67	.229	.361	.134	3.068	.387
541	68	.471	.618	.309	.807	.786
542	65	.036	.047	.338	8.892	.047
543	70	.212	.348	.114	3.464	.371
544	69	.223	.285	.638	-1.239	.297

ITEM	N	POINT BISERIAL	BISERIAL R	PROPORTION CORRECT	DIFFICULTY (B)	DISCRIMINATION (A)
545	69	.293	.417	.203	1.993	.459
546	61	.320	.427	.721	-1.372	.472
547	69	.187	.236	.580	-.855	.243
548	63	-.233	-.456	.063	-3.355	-.512
549	63	.253	.319	.556	-.443	.335
550	70	.120	.151	.443	.949	.153
551	68	.416	.522	.529	-.139	.612
552	67	.344	.432	.537	-.215	.479
553	66	-.179	-.267	.167	-3.618	-.277
554	68	.086	.116	.265	5.414	.117
555	70	.015	.025	.114	48.221	.025
556	66	-.152	-.232	.152	-4.431	-.239
557	67	.028	.035	.493	.501	.035
558	56	.053	.078	.172	12.132	.078
559	66	.317	.485	.848	-2.119	.555
560	69	.259	.411	.130	2.741	.451
561	66	.287	.381	.288	1.468	.412
562	67	.362	.477	.299	1.105	.543
563	67	.059	.082	.224	9.253	.082
564	66	.216	.304	.212	2.630	.319
565	64	.046	.064	.234	11.340	.064
566	64	-.034	-.043	.531	1.809	-.043
567	64	.259	.404	.141	2.663	.442
568	69	.528	.662	.522	-.083	.883
569	63	.167	.215	.349	1.805	.220
570	65	.168	.256	.846	-3.982	.265
571	63	.109	.181	.111	6.747	.184
572	66	.289	.380	.303	1.357	.411
573	65	.377	.523	.231	1.406	.614
574	64	.036	.047	.328	9.478	.047
575	69	-.088	-.125	.203	-6.648	-.126
576	67	.228	.286	.463	.325	.298
577	66	.207	.270	.318	1.753	.280
578	67	.030	.038	.507	-.462	.038
579	64	.149	.274	.078	5.178	.285
580	70	.145	.185	.371	1.779	.188
581	68	.338	.437	.338	.956	.486
582	66	.473	.593	.515	-.063	.736
583	68	.530	.706	.309	.706	.997
584	66	.396	.533	.727	-1.133	.630
585	64	.022	.031	.203	26.805	.031
586	68	.317	.415	.309	1.202	.456
587	66	.295	.389	.348	1.028	.411
588	71	.279	.361	.662	-1.158	.387
589	67	.174	.222	.388	1.232	.228
590	66	.290	.383	.508	-.055	.390
591	66	.172	.237	.235	3.048	.244
592	70	.329	.423	.357	.866	.467
593	66	.301	.377	.515	-.100	.407
594	66	.084	.109	.353	3.461	.110
595	67	.355	.483	.254	1.371	.552
596	70	-.063	-.119	.371	-12.339	-.120
597	65	.345	.458	.600	-.578	.487
598	66	.361	.460	.379	.670	.518
599	70	.543	.683	.557	-.210	.935

ITEM	N	POINT BISERIAL	BISERIAL R	PROPORTION CORRECT	DIFFICULTY (B)	DISCRIMINATION(A)
001	68	.030	.038	.397	6.872	.038
001	69	.363	.468	.353	.806	.530
002	70	.160	.222	.771	-3.343	.228
003	66	.198	.238	.409	.967	.245
004	69	.138	.196	.203	4.240	.200
005	70	.104	.177	.100	7.240	.180
006	69	.253	.333	.696	-1.540	.353
007	63	.404	.528	.317	.902	.622
008	65	-.043	-.066	.154	-15.446	-.066
009	69	.249	.349	.217	2.242	.372
010	72	.262	.340	.333	1.270	.362
011	67	-.054	-.085	.134	-13.032	-.085
012	64	.052	.090	.094	14.628	.090
013	61	-.031	-.049	.131	-22.891	-.049
014	68	.269	.477	.088	2.837	.543
015	65	.266	.353	.708	-1.551	.377
016	64	.230	.297	.344	1.352	.311
017	65	.354	.465	.308	1.079	.525
018	66	.060	.082	.258	7.921	.082
019	67	-.140	-.220	.134	-5.035	-.226
020	66	.117	.184	.136	5.970	.187
021	75	.011	.014	.440	10.784	.014
022	68	.261	.329	.574	-.567	.348
023	63	.111	.178	.127	6.408	.181
024	67	-.194	-.328	.104	-3.839	-.347
025	65	.089	.111	.477	.520	.112
026	144	.432	.544	.563	-.292	.648
027	144	.252	.384	.153	2.666	.416
028	144	.273	.461	.896	-2.731	.519
029	143	.296	.371	.538	-.257	.400
030	143	.633	.794	.517	-.054	1.306
031	142	.165	.212	.359	1.703	.217
032	141	.135	.186	.766	-3.902	.189
033	140	.357	.448	.514	-.078	.501
034	140	.271	.341	.550	-.369	.363
035	140	.316	.398	.557	-.360	.434
036	140	.378	.474	.529	-.153	.538
037	140	.419	.599	.300	-1.405	.748
038	140	.193	.245	.607	-1.108	.253
039	140	.230	.297	.657	-1.361	.311
040	140	.392	.555	.793	-1.472	.667
041	140	.318	.458	.807	-1.893	.515
042	140	.281	.386	.757	-1.805	.418
043	140	.264	.405	.850	-2.559	.443
044	140	.316	.396	.507	-.044	.431
045	140	.356	.448	.557	-.320	.501
046	140	.371	.495	.721	-1.183	.570
047	140	.216	.290	.271	2.103	.303
048	139	.343	.457	.719	-1.269	.514
049	139	.277	.402	.813	-2.211	.439
050	139	.264	.373	.791	-2.171	.402
051	138	.387	.489	.406	.436	.561
052	138	.377	.515	.246	1.334	.601
053	138	.302	.410	.746	-1.615	.450
054	137	.137	.172	.482	.262	.175

ITEM	N	POINT BISERIAL	BISERIAL R	PROPORTION CORRECT	DIFFICULTY (B)	DISCRIMINATION(A)
055	137	.288	.361	.489	.076	.387
056	137	.318	.405	.387	.709	.443
057	137	.345	.457	.708	-1.198	.514
058	137	.020	.032	.131	35.052	.032
059	137	.239	.332	.226	2.265	.352
060	136	.291	.375	.353	1.006	.405
061	136	.398	.501	.559	-.296	.579
062	136	.334	.497	.169	1.928	.573
063	136	.196	.291	.169	3.293	.304
064	136	.255	.643	.029	2.948	.840
065	136	.491	.670	.250	1.007	.903
066	136	.382	.482	.419	.424	.550
067	136	.385	.494	.640	-.726	.568
068	136	.277	.361	.324	1.265	.387
069	136	.221	.302	.757	-2.307	.317
070	136	.419	.527	.559	-.282	.620
071	136	.346	.464	.728	-1.308	.524
072	136	.390	.648	.110	1.893	.851

Appendix B

Cross-tabulation of the 369 items selected for the final item pool by traditional difficulty and discrimination indices (cell entries are item reference numbers)

Item Discrimination (Item-Test Biserial r)	Item difficulty: proportion correct (p)										
	0- .099	.100- .199	.200- .299	.300- .399	.400- .499	.500- .599	.600- .699	.700- .799	.800- .899	.900- .999	
1.000 to .900										25 102 42 71 28 7 64	1.000 to .900
.899 to .800									191 194	14 196 11 84 24 13 138 190	.899 to .800
.799 to .700				337 583	347	630 380 272		87	129 51 44 40 101 36	122 206 99 22 124 125 9 27 70 96	.799 to .700
.699 to .600	664	672	665	541 114	283 340 296 342 266 265 264 301 315	329 161 56 568 599	270 332 91 130 128 503 143 522	43 85 109 501 239 515	127 86 186 103 83 173 90 47 199	8 126 65 158 181 80 105 134 66	.699 to .600
.599 to .500		379 166 398	120 652 573 254 174 359 319	294 321 607	386 538 60 59 526 113	104 670 551 146 582 52 626 145 209 661 144	188 33 313 382 46 390 37 211 154 207 365 183 203 208 221 110 156 535	285 123 640 349 322 149 204 584	106 88 34 189 311 232 227	23 89 19 184 182 5 32 31 16 63 121 255 17 262 198	.599 to .500

--continued next page--

Appendix B, continued

Item difficulty: proportion correct (p)

Item Discrimination (Item-Test Biserial, r)	0-.099	.100-.199	.200-.299	.300-.399	.400-.499	.500-.599	.600-.699	.700-.799	.800-.899	.900-.999	Item Discrimination (Item-Test Biserial r)
.499 to .400	614	662 328 560 567 336 309	288 545 152 562 162 595 140 378 263 273 381 115	299 215 601 231 617 656 526 291 581 610 586 521 592 525 598 572 306 587 606	651 666 271 302 506 375 111 267	327 633 369 391 636 295 50 233 502 519 552 549 645 590 593 622	667 597 307 393 224 58 222 292 53 205 234 287	108 646 241 185 94 648 653 657 235 117 671 293 112 546	637 643 559 95 82 649 93 641 76	18 69 212 131 2 187 240 202 100 628	.499 to .400
.399 to .295	290 527 537	298 303 388 540 385 543 627 237 312 364	180 534 271 561 297 564 253 383 245 659 609 193 514 360 528	217 168 304 216 660 159 668 107 147 400 367 260 397 286 341	238 139 164 655 324 148	635 629 377 355 644 116 133 634 258 289 318 252	137 588 155 176 218 157 639	650 214 276 141 642 669 371 26	135 151 73 81 256 507	20 261 4 74 62 30 78 132 77 12	.399 to .295

Matrices of tetrachoric interitem correlations and principal factor loading tables for the six real and six random data subtests

TABLE C-1

Tetrachoric Item Intercorrelations for
Test 1 (upper triangle) and Test 2 (lower triangle)

Items (test 2)		Items (test 1)																			Items (test 1)
		211	203	202	190	180	168	166	159	157	134	131	94	90	59	58	47	46	43	36	
		37	41	24	04	19	11	40	01	03	22	17	21	07	20	08	25	20	33	18	233
56	33		25	27	35	36	21	42	09	28	45	-14	27	30	36	27	32	35	32	34	211
66	37	43		32	40	43	18	48	13	22	32	27	04	25	20	24	33	36	34	40	203
84	32	44	76		30	26	-00	27	27	-02	35	20	04	42	44	11	43	38	34	35	202
106	38	20	42	60		89	34	18	03	26	73	31	44	63	40	14	45	15	43	72	190
107	-07	20	14	30	18		25	48	26	16	30	14	36	36	39	04	36	37	14	37	180
109	15	32	45	49	44	29		38	07	-08	05	-12	35	44	35	11	07	18	26	14	168
115	10	37	36	12	28	26	26		41	43	87	16	25	90	35	21	44	60	46	30	166
123	22	24	21	02	14	19	48	18		27	07	34	14	17	36	-05	05	21	08	22	159
127	21	32	61	63	43	11	56	23	24		26	13	13	12	18	21	34	29	12	23	157
1-	32	12	15	07	16	12	32	40	47	43		02	60	62	24	22	35	03	12	43	134
152	10	33	06	12	06	07	13	05	30	23	36		03	33	37	-01	32	19	47	39	131
162	15	22	21	13	22	21	23	32	12	39	31	23		30	18	21	23	13	20	31	94
173	26	34	23	52	10	20	31	30	40	37	40	10	19		31	07	29	20	20	64	90
174	18	29	43	20	36	15	18	38	35	13	18	24	30	38		18	42	31	59	39	59
185	23	44	03	24	07	18	01	29	31	12	16	35	33	42	14		36	17	37	12	58
189	40	21	30	39	25	00	45	30	24	50	14	23	16	30	32	30		33	63	52	47
193	23	19	19	-14	-12	-07	13	-01	42	49	17	26	04	10	20	23	30		38	31	46
231	14	24	28	33	06	-05	15	33	20	22	14	02	33	25	40	28	05	11		50	43
237	87	32	-09	-32	19	-10	-06	20	35	05	33	20	35	25	28	34	-05	11	19		
	31	56	66	84	106	107	109	115	123	127	140	152	162	173	174	185	189	193	231		

Items (test 2)

Note: Decimal Points Omitted.

TABLE C-2

Tetrachoric Item Intercorrelations for
Test 3 (upper triangle) and Test 4 (lower triangle)

Items (test 4)		Items (test 3)																			Items (test 3)
		214	207	191	148	145	132	130	117	116	93	87	86	85	81	80	63	53	39	13	
		39	22	54	04	05	32	40	30	14	04	42	39	34	24	13	19	10	08	41	235
34	25		26	27	24	13	05	38	29	15	23	40	19	37	11	23	12	18	09	00	214
44	21	40		42	09	22	26	23	41	40	22	32	20	39	12	30	22	17	-12	23	207
50	12	-01	35		05	26	30	48	51	43	25	56	44	42	41	31	36	27	24	71	191
52	27	22	13	33		32	-05	13	-01	16	38	24	24	14	-09	22	-00	16	11	02	148
76	18	56	46	11	30		16	40	29	28	19	33	36	42	08	27	35	22	23	39	145
82	13	36	06	-05	20	35		51	07	27	17	33	31	17	08	17	-19	02	13	57	132
101	17	69	58	32	37	60	55		55	36	04	45	62	42	32	30	35	03	24	43	130
104	14	31	54	39	25	30	29	39		22	0	38	40	24	31	22	23	12	-36	20	117
114	13	24	50	21	32	15	02	13	32		17	26	29	52	16	25	10	04	50	37	116
143	-01	15	36	29	32	17	17	35	44	36		33	-01	10	12	12	22	40	08	32	93
147	28	-06	38	22	22	26	01	27	08	16	26		38	53	27	52	55	35	31	63	87
149	09	42	27	21	25	22	26	23	24	48	36	31		47	15	18	46	13	36	50	86
158	42	12	43	46	44	05	-05	55	27	25	34	41	08		23	24	41	25	37	49	85
164	10	12	31	34	-06	07	26	29	26	38	12	-02	19	32		34	24	14	00	24	81
183	13	15	31	26	23	25	37	40	33	14	31	16	16	06	12		50	39	60	45	80
221	10	46	56	16	49	30	18	48	34	22	40	23	58	24	06	30		18	42	14	63
227	45	24	33	31	28	20	22	35	52	22	38	-08	34	51	20	11	21		22	24	53
234	15	-01	15	11	43	03	00	18	23	20	16	25	21	32	16	-06	22	26		40	39
238	04	03	29	31	22	17	33	12	23	30	24	12	13	04	31	02	25	18	32		
	26	34	44	50	52	76	82	101	104	114	143	147	149	158	164	183	221	227	234		

Items (test 4)

Note: Decimal Points Omitted.

TABLE C-3

Tetrachoric Item Intercorrelations for
Test 5 (upper triangle) and Test 6 (lower triangle)

Items (test 6)	Items (test 5)																			Items (test 5)
	217	209	208	199	194	188	161	154	151	133	128	112	111	110	103	95	91	69	60	
	04	30	21	21	42	32	34	12	-06	22	24	35	37	21	31	27	13	30	22	224
37	20	29	23	27	03	15	38	33	23	30	33	25	20	17	30	16	33	01	18	217
83	28	09	12	39	42	39	35	12	-08	11	38	18	45	18	42	06	22	10	47	209
108	29	26	25	13	45	34	36	42	13	06	35	20	16	50	14	19	40	04	12	208
113	42	38	27	09	49	32	58	35	35	26	28	40	-04	20	36	44	39	56	37	199
125	61	20	28	38	30	35	58	57	24	14	34	30	16	26	46	54	51	48	37	194
135	35	22	38	15	23	46	42	28	06	34	35	12	30	32	50	08	27	16	33	188
137	04	17	01	08	09	04	34	45	35	20	36	27	18	31	24	17	52	19	35	161
139	05	03	34	26	33	25	-05	20	11	20	29	02	12	15	12	43	28	39	22	154
141	27	23	10	10	16	27	05	13	17	14	34	21	15	17	08	-06	28	-04	05	151
156	27	28	24	15	35	27	26	23	42	32	49	13	19	20	30	20	27	01	33	133
176	33	05	43	19	16	24	10	04	00	-10	08	40	24	26	51	12	34	23	44	128
184	46	20	23	22	16	17	30	30	22	29	10	15	26	24	49	25	41	16	15	112
204	36	33	-06	19	31	26	12	28	10	21	45	14	20	12	41	28	24	15	35	111
205	32	15	14	18	34	04	09	-09	22	23	34	-17	-12	25	36	04	35	25	25	110
215	37	45	35	06	13	37	11	24	18	05	44	-00	-02	06	15	23	48	32	25	103
216	13	17	07	13	16	51	41	24	11	12	24	04	09	09	30	-01	36	48	20	95
222	25	32	38	22	44	06	08	02	20	18	02	16	42	12	-04	21	11	30	47	91
232	36	40	13	28	30	15	23	18	18	15	24	01	30	25	41	18	32	29	32	69
239	35	28	15	18	28	25	-05	26	28	10	40	16	28	52	15	57	24	11	18	
	27	37	83	108	113	125	135	137	139	141	156	176	184	204	205	215	216	222	232	
Items (test 6)																				

Note: Decimal Points Omitted.

TABLE C-4

Tetrachoric Item Intercorrelations for
Test 7 (upper triangle) and Test 8 (lower triangle)

Items (test 8)	Items (test 7)																			Items (test 7)
		19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	
1		08	-17	-09	-10	-08	-12	-01	-14	-14	-04	01	-28	-12	03	-05	05	06	-03	-21	20
2	08		-02	-15	16	-07	-13	-10	03	-10	03	-10	18	15	05	02	-11	-02	-06	10	19
3	06	-11		-06	-14	-21	14	15	-05	-11	08	-02	00	-13	-07	14	-00	-13	-05	08	18
4	-15	-13	05		04	14	-21	14	14	04	-05	09	-11	-14	-23	09	-02	-16	-01	21	17
5	-07	05	-05	-12		-04	-06	-14	02	12	-14	-09	16	-03	-05	00	-00	09	08	13	16
6	00	-08	-06	02	-03		10	12	16	09	19	-06	07	-02	-02	-03	-07	-11	-03	16	15
7	-13	-03	07	05	08	20		17	10	07	07	-04	09	-08	-02	06	11	-12	-00	-00	14
8	16	-04	12	15	-14	-06	-12		06	-11	-04	-03	-10	17	-08	14	10	19	17	16	13
9	-38	-08	08	28	04	-13	-15	-03		09	06	20	10	01	05	07	-00	-11	-16	00	12
10	-16	-11	18	15	-04	03	-05	-17	17		-17	20	-13	17	05	-19	-04	-27	-02	03	11
11	20	-05	00	05	00	-00	03	00	-13	-23		-13	34	11	08	01	-11	-08	-10	03	10
12	15	07	11	10	-15	-08	12	01	05	-02	02		-01	-08	05	16	11	-12	-14	06	9
13	02	22	-07	26	-04	05	11	-21	-08	-08	08	00		02	17	05	02	-13	-00	05	8
14	13	-05	-15	13	04	26	-02	-22	-01	-16	-00	-05	-08		-03	-05	01	11	-17	05	7
15	-10	-15	14	02	07	03	06	13	-00	07	-07	15	08	-00		04	03	13	-08	-02	6
16	-06	-16	-10	02	12	-04	-03	04	-04	-08	10	-14	07	-15	-04		-02	18	-10	25	5
17	10	01	08	-02	-22	10	02	07	13	-03	03	-02	-08	-17	16	-37		-01	-06	-02	4
18	08	16	12	07	-00	-05	09	-00	-03	-07	18	10	17	00	-08	-04	04		11	02	3
19	08	-16	-01	-03	10	05	-09	04	09	-03	08	-07	04	-00	05	01	06	-01		-29	2
20	10	11	04	-08	00	-00	-07	-06	-00	-06	16	-02	-02	06	-20	-17	13	24	08		1
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19		
	Items (test 8)																				

Note: Decimal Points Omitted.

TABLE C-5

Tetrachoric Item Intercorrelations for
Test 9 (upper triangle) and Test 10 (lower triangle)

Items (test 10)	Items (test 9)																			Items (test 9)
	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	
1	-11	-13	03	17	05	-09	05	-08	05	-06	10	-09	07	02	25	-08	08	18	03	20
2	09	05	-16	05	08	05	06	06	02	09	0	03	-12	10	00	-05	-13	-01	06	19
3	02	-05	-01	-09	-13	03	01	08	20	16	-15	17	06	15	06	-03	08	11	28	18
4	20	-10	05	16	-02	-07	11	08	-02	-14	-06	-30	06	-23	11	-02	15	02	10	17
5	-01	08	-03	05	08	-21	-02	04	02	-09	0	-16	06	06	03	-22	-07	-12	07	16
6	05	-02	06	08	13	02	10	-03	13	-06	-24	00	-15	-09	-10	-07	-10	-07	-02	15
7	-01	-00	08	-16	-11	12	01	-12	-05	20	-03	-07	-00	12	07	02	06	08	-10	14
8	04	-02	-13	-02	10	-13	-02	-10	16	08	-05	-19	18	-01	-17	-13	-10	-26	-27	13
9	-04	10	-21	-20	05	18	-00	-15	-16	01	-11	01	-05	-14	03	-00	00	06	-11	12
10	-13	06	-02	06	12	12	-03	-12	-00	-16	-08	-16	08	05	-10	-07	-06	-00	05	11
11	-17	09	-21	03	02	08	03	24	29	03	-02	20	-04	08	-01	04	-02	-03	-12	10
12	-07	-14	06	02	-11	-04	06	-05	02	-02	-01	08	-10	0	-05	-18	02	11	-13	9
13	-07	13	02	-10	-11	02	10	-02	03	-10	03	-01	-14	-21	-12	-03	-18	14	-12	8
14	-05	-05	03	-05	16	-00	-05	17	05	08	-08	-07	12	01	-09	02	05	-21	-10	7
15	07	-06	04	15	04	13	-19	02	01	-21	01	01	21	07	00	02	13	-04	-07	6
16	13	07	-02	-03	02	02	07	-11	-10	00	-10	11	-03	-05	02	16	-02	09	15	5
17	19	-11	-04	-05	-14	03	-21	04	-11	21	-14	-08	05	-04	-09	18	-20	03	05	4
18	-15	-22	18	-06	-02	14	-02	-04	01	23	-22	06	-16	-22	-13	-00	-06	-13	-02	3
19	12	18	-04	08	-20	-10	-11	04	-08	21	-08	09	-08	-13	-02	11	06	24	-02	2
20	-04	06	08	10	-11	08	03	-16	-00	-03	16	16	-00	05	14	-20	-04	01	-08	1
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
	Items (test 10)																			

Note: Decimal Points Omitted.

TABLE C-6

Tetrachoric Item Intercorrelations for
Test 11 (upper triangle) and Test 12 (lower triangle)

Items (test 12)	Items (test 11)																			Items (Test 11)
	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	
1		-15	03	12	-02	04	03	22	-00	28	-05	05	22	-08	-18	-08	06	-14	25	-04	20
2	-03		-02	00	-00	25	-12	08	08	04	06	-00	-02	-13	14	-16	-08	08	07	-09	19
3	08	-05		-02	07	-05	03	-20	-11	34	04	04	-10	-05	10	-18	13	-04	08	03	18
4	10	-03	-12		11	-13	05	-05	03	06	-25	24	18	03	-05	16	-05	-11	-03	-05	17
5	07	17	05	-04		-06	-03	-08	05	-06	03	10	01	08	10	-00	04	-01	03	05	16
6	-16	01	01	-17	-07		-18	08	-18	14	20	-10	-25	-10	04	-02	04	11	04	-02	15
7	-10	02	15	-06	-06	08		-04	12	-02	00	-10	23	-15	-07	02	-16	-04	12	06	14
8	11	-02	16	28	-08	02	02		03	08	08	12	19	02	-01	06	-17	-24	-24	-14	13
9	12	18	-10	08	11	-03	-04	-13		03	13	-11	09	-13	04	-05	06	-17	12	09	12
10	13	19	05	-16	24	-09	-11	-08	25		08	-02	02	-10	-01	-07	15	09	-04	08	11
11	-06	-12	04	19	09	-18	14	08	04	01		-03	-12	08	16	-20	14	12	00	11	10
12	06	09	15	10	04	05	22	08	-01	06	11		14	01	-00	-10	-02	05	-17	-08	9
13	05	02	-00	-02	-05	08	02	-03	16	12	01	15		02	06	-08	-04	13	09	03	8
14	-22	-12	-03	02	-01	10	20	03	04	07	-00	13	10		-15	-03	12	12	-07	12	7
15	-03	-04	05	-22	00	20	-01	-15	-01	06	-06	06	-02	01		-00	-17	-12	03	05	6
16	06	-26	-11	-16	-29	-09	-00	-02	-18	-14	07	-04	-05	-02	13		-08	-04	-07	-24	5
17	05	-12	-19	12	12	-04	-02	0	-03	01	15	28	00	-07	05	08		03	-01	29	4
18	06	-13	-05	-19	13	17	-14	-05	05	-07	04	-20	08	08	22	-03	-02		-01	09	3
19	02	-08	-03	-05	-05	05	-15	-07	04	-02	-08	08	16	13	-02	-05	-07	-05		-20	2
20	-05	09	13	01	-09	-12	-04	0	16	09	-09	-04	-04	08	-11	08	04	-11	-13		1
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19		
	Items (test 12)																				

Note: Decimal Points Omitted.

TABLE C-7

Principal Factor Loadings Matrix for Test 1

Item	Factor									
	1	2	3	4	5	6	7	8	9	h^2
36	-.71	.06	.33	-.04	-.22	.03	-.00	-.05	-.06	.68
43	-.61	.47	.06	.25	-.16	-.25	.03	-.02	.04	.75
46	-.51	.34	-.26	-.05	.29	.06	.01	-.10	-.16	.57
47	-.65	.28	.05	.26	-.20	.11	.00	-.02	-.11	.64
58	-.28	.07	-.17	.45	-.12	-.02	.22	-.15	-.01	.40
59	-.60	.30	.11	-.02	.09	-.22	.13	.27	-.11	.63
90	-.73	-.34	-.01	-.44	-.23	-.25	-.00	-.20	-.10	1.01
94	-.46	-.38	.03	.25	.03	-.13	.19	.28	.24	.60
131	-.36	.42	.28	-.28	-.21	.04	.10	-.03	.27	.59
134	-.70	-.56	-.19	.05	-.31	.18	-.10	.18	.03	1.01
157	-.36	.06	-.20	.05	-.03	.41	.36	-.04	-.06	.48
159	-.31	.18	-.08	-.41	.16	.11	.28	.27	.10	.49
166	-.81	-.11	-.59	-.26	-.02	-.02	.05	-.12	.06	1.11
168	-.35	-.16	-.04	.04	.27	-.55	.12	-.12	.01	.55
180	-.66	-.18	.28	-.02	.58	.21	-.05	-.06	.06	.95
190	-.78	-.32	.57	.12	.05	.10	.00	-.09	-.03	1.07
202	-.49	.20	.06	-.22	-.04	.02	-.35	.23	-.20	.51
203	-.54	.16	-.07	.04	.09	.16	-.23	-.25	.20	.52
211	-.54	-.07	-.22	.26	.13	.01	-.10	.14	-.21	.50
233	-.36	.17	-.24	.16	.05	-.05	-.35	.10	.30	.48
Factor Contribution	6.37	1.58	1.26	1.07	.88	.80	.66	.52	.43	13.57
% of Common Variance	52.50	13.07	10.40	8.81	7.24	6.59	5.41	4.30	3.55	111.87
% of Total Variance	31.83	7.93	6.31	5.34	4.39	4.00	3.28	2.61	2.15	67.83

TABLE C-8

Principal Factor Loadings Matrix for Test 2

Item	Factor									
	1	2	3	4	5	6	7	8	9	h^2
31	-.57	.42	.57	-.36	.05	.10	-.08	-.01	-.02	.97
56	-.58	.05	.03	.14	.26	.11	.07	.13	.20	.50
66	-.67	-.43	.18	-.10	.10	-.30	-.03	.04	.20	.82
84	-.65	-.62	.24	.04	.20	.21	-.04	-.12	.05	.97
106	-.53	-.25	.40	.01	-.29	.03	.13	.18	-.06	.63
107	-.26	-.18	-.13	.30	-.17	.23	.00	.13	.14	.33
109	-.60	-.31	-.15	-.12	-.31	.07	-.08	.05	.02	.61
115	-.50	.04	-.04	.36	-.12	-.17	.02	.05	-.17	.45
123	-.52	.27	-.34	-.08	-.18	.02	-.27	.12	.12	.60
127	-.70	-.27	-.14	-.31	-.05	-.06	.21	-.24	.01	.78
140	-.50	.25	-.21	-.02	-.36	.03	.00	-.24	.00	.54
152	-.35	.21	-.28	-.05	.07	.16	.26	.14	.10	.38
162	-.45	.15	-.04	.26	-.09	-.11	.34	-.13	-.06	.38
173	-.59	.05	-.11	.18	.08	.20	-.34	-.21	-.07	.60
174	-.52	.09	-.02	.21	.02	-.30	-.12	.26	-.07	.51
185	-.44	.29	-.15	.23	.34	.26	.09	.01	-.07	.55
189	-.54	-.12	-.10	-.28	.12	.08	.00	.15	-.41	.60
193	-.32	.22	-.40	-.46	.20	-.18	.02	.06	.07	.59
231	-.40	.05	.00	.27	.22	-.35	-.07	-.19	-.01	.45
237	-.38	.85	.36	.01	-.10	-.00	.03	-.03	.07	1.02
Factor Contribution	5.34	2.10	1.21	1.05	.75	.65	.48	.43	.36	12.36
% of Common Variance	50.24	19.80	11.36	9.87	7.07	6.12	4.53	4.03	3.37	116.40
% of Total Variance	26.68	10.51	6.03	5.24	3.75	3.25	2.41	2.15	1.79	61.82

TABLE C-9

Principal Factor Loadings Matrix for Test 3

Item	Factor									
	1	2	3	4	5	6	7	8	9	h^2
13	-.74	.03	.40	-.22	.20	.01	-.17	.13	-.01	.84
39	-.44	.64	.33	.22	.02	.02	.16	-.04	.05	.80
53	-.36	.29	-.26	-.27	.13	.05	-.06	.05	.01	.37
63	-.53	.26	-.32	.40	.17	.03	-.08	.12	-.11	.67
80	-.58	.39	-.12	.05	.25	-.08	-.02	-.36	-.06	.70
81	-.38	-.15	-.13	.06	.36	-.11	.01	-.11	.26	.41
85	-.67	.06	.01	.12	-.18	-.11	.19	.19	-.09	.59
86	-.65	-.09	.14	.32	-.22	.19	-.16	.14	.03	.68
87	-.77	.09	-.14	-.10	.12	.17	-.00	-.01	-.16	.70
93	-.32	.24	-.20	-.50	-.04	.04	-.07	.09	.09	.47
116	-.53	.09	.23	-.01	-.24	-.42	.22	-.00	.11	.64
117	-.49	-.53	-.34	.10	-.01	-.14	-.16	-.07	.02	.69
130	-.70	-.28	.11	.22	-.16	.10	-.10	-.23	.08	.68
132	-.42	-.22	.52	-.29	-.03	.07	-.09	-.21	-.12	.67
145	-.50	.14	-.07	.05	-.27	-.09	-.31	.04	.00	.46
148	-.23	.25	-.21	-.21	-.44	.16	-.07	-.06	.15	.44
191	-.76	-.21	.02	-.10	.23	-.08	.06	.20	.15	.76
207	-.47	-.19	-.19	-.19	-.10	-.32	.08	-.01	-.22	.50
214	-.41	-.10	-.31	-.07	-.20	.20	.34	-.15	.00	.49
235	-.52	-.33	.06	-.02	.11	.33	.27	.09	-.00	.58
Factor Contribution	5.92	1.51	1.18	.96	.83	.60	.53	.41	.26	12.20
% of Common Variance	56.04	14.32	11.19	9.07	7.89	5.69	5.00	3.91	2.47	115.59
% of Total Variance	29.62	7.57	5.92	4.79	4.17	3.01	2.64	2.06	1.31	61.09

TABLE C-10

Principal Factor Loadings Matrix for Test 4

Item	Factor									
	1	2	3	4	5	6	7	8	9	h^2
26	-.36	-.13	.43	.01	.21	.13	-.08	.30	.17	.51
34	-.56	.56	.10	-.06	.15	.32	-.02	-.10	-.08	.78
44	-.72	-.02	-.05	.00	-.35	.24	-.08	-.10	.13	.72
50	-.47	-.36	-.01	.19	-.22	-.11	.07	-.02	-.05	.46
52	-.54	-.15	.12	-.27	.24	-.28	.07	-.09	.02	.56
76	-.54	.44	.14	-.08	-.10	-.02	-.17	-.07	.26	.62
82	-.41	.47	-.06	.26	.21	-.34	-.13	.18	-.06	.67
101	-.76	.31	.29	.08	-.14	-.09	-.08	-.12	-.21	.87
104	-.63	-.03	-.11	.27	-.02	.00	.23	-.10	.20	.58
114	-.50	-.21	-.37	-.02	.02	.26	-.10	.09	.06	.52
143	-.55	-.14	-.21	-.04	-.08	-.12	.27	.00	-.02	.46
147	-.37	-.22	.11	-.39	-.27	-.11	-.21	.22	.04	.53
149	-.54	.06	-.36	-.28	.19	.17	.06	.24	-.14	.66
158	-.56	-.45	.41	.03	-.06	.07	-.00	-.03	-.19	.74
164	-.37	-.14	-.17	.42	-.04	.10	-.27	.06	-.19	.49
183	-.41	.20	-.03	.07	-.25	-.26	.21	.23	.03	.45
221	-.64	.14	-.17	-.37	.02	-.01	.13	-.12	-.10	.63
227	-.56	-.18	.16	.28	.32	.13	.24	.03	.08	.64
234	-.35	-.36	-.02	-.16	.30	-.15	-.21	-.18	.02	.46
238	-.37	-.13	-.31	.15	.10	-.20	-.28	-.11	.12	.43
Factor Contribution	5.50	1.56	1.00	.95	.76	.69	.58	.42	.34	11.80
% of Common Variance	54.17	15.43	9.84	9.32	7.44	6.77	5.69	4.11	3.39	116.16
% of Total Variance	27.48	7.83	4.99	4.73	3.78	3.43	2.88	2.09	1.72	58.92

TABLE C-11

Principal Factor Loadings Matrix for Test 5

Item	Factor									
	1	2	3	4	5	6	7	8	9	h^2
60	-.56	-.09	.16	.02	.28	.08	-.08	.28	.20	.56
69	-.49	.49	.26	.10	-.04	.05	-.23	.09	-.02	.62
91	-.66	.01	-.16	.10	-.13	-.06	.02	.28	.06	.58
95	-.48	.48	.19	.10	-.02	-.35	.07	.00	-.01	.63
103	-.64	-.24	.24	.12	-.18	-.01	-.09	-.01	-.19	.62
110	-.47	-.14	-.15	-.24	-.28	.09	-.30	.14	-.05	.52
111	-.43	-.28	.33	-.08	-.01	-.21	.26	.09	.11	.51
112	-.50	-.13	.06	.30	-.43	-.04	.11	-.07	-.04	.56
128	-.62	-.30	-.08	.10	.06	-.10	-.13	-.06	.08	.54
133	-.42	-.27	-.04	.15	.26	-.28	-.28	-.17	.09	.54
151	-.29	-.12	-.44	.26	-.08	.08	.08	-.05	.20	.42
154	-.53	.38	-.25	-.21	.26	-.20	.05	-.05	-.09	.65
161	-.68	.04	-.29	-.05	.10	.24	.20	-.10	.05	.67
188	-.56	-.22	.08	-.27	.13	.12	-.14	-.16	-.10	.53
194	-.74	.34	.02	-.14	-.06	.10	.08	-.06	.11	.71
199	-.65	.28	-.08	.34	.09	.27	-.05	-.08	-.10	.72
208	-.48	-.00	-.33	-.45	-.22	-.12	.03	.04	-.02	.60
209	-.52	-.25	.30	-.09	.22	.24	.20	.09	-.11	.59
217	-.42	-.21	-.25	.16	.15	-.19	.17	.06	-.26	.47
224	-.47	-.01	.34	-.15	-.14	.00	.06	-.28	.13	.48
Factor Contribution	5.85	1.32	1.09	.81	.71	.59	.49	.37	.29	11.52
% of Common Variance	58.67	13.24	10.96	8.10	7.13	5.94	4.89	3.75	2.93	115.60
% of Total Variance	29.23	6.60	5.46	4.03	3.55	2.96	2.44	1.87	1.46	57.61

TABLE C-12

Principal Factor Loadings Matrix for Test 6 (Random data)

Items	Factor									
	1	2	3	4	5	6	7	8	9	h^2
27	-.68	-.21	.02	.07	.03	.40	.07	-.18	.10	.72
37	-.52	.14	-.08	-.17	-.10	-.01	.38	.07	-.08	.49
83	-.45	-.38	-.20	.14	.30	-.21	-.04	.06	.10	.56
108	-.42	-.16	.02	-.06	.07	.06	-.13	-.02	-.37	.37
113	-.57	.02	-.07	-.18	.21	.07	.00	.20	.20	.49
125	-.62	-.22	.28	.37	.03	.11	-.04	-.18	-.11	.71
135	-.45	-.29	.42	.09	-.12	-.18	.15	.07	.21	.59
137	-.32	.10	.06	-.00	-.45	-.33	-.06	.06	.06	.44
139	-.41	.11	-.17	-.04	.22	-.31	-.40	-.05	-.07	.53
141	-.36	.10	.07	-.21	-.00	.03	-.12	-.36	.13	.35
156	-.59	.34	-.00	.15	.05	-.11	-.14	.03	.19	.56
176	-.26	-.44	-.19	.20	-.04	.18	-.13	.30	.01	.49
184	-.46	-.32	-.10	-.31	-.33	-.05	-.12	-.19	.02	.57
204	-.51	.30	-.03	-.07	-.30	.32	-.16	.16	.07	.60
205	-.37	.37	.23	-.12	.42	.11	-.02	.04	.02	.53
215	-.50	.21	-.33	.38	.05	-.17	.36	-.15	-.02	.73
216	-.42	.01	.53	.06	-.02	-.18	-.02	.14	-.17	.54
222	-.41	-.28	-.30	-.38	.08	-.10	.13	.03	-.04	.52
232	-.52	.09	.14	-.33	.07	-.01	.16	.07	-.15	.46
239	-.57	.31	-.30	.22	-.22	.07	-.08	.06	-.14	.64
Factor Contribution	4.62	1.27	1.04	.91	.85	.70	.63	.45	.40	10.89
% of Common Variance	50.15	13.79	11.25	9.90	9.24	7.65	6.83	4.92	4.38	118.10
% of Total Variance	23.12	6.36	5.18	4.56	4.26	3.53	3.15	2.27	2.02	54.44

TABLE C-13

Principal Factor Loadings Matrix for Test 7 (Random data)

Item	Factor									
	1	2	3	4	5	6	7	8	9	h^2
1	.44	-.10	-.24	.31	-.02	.02	-.00	.12	-.14	.39
2	-.33	.02	-.04	-.19	-.38	.17	-.08	-.19	.10	.38
3	-.26	.21	-.34	.28	-.04	-.08	-.21	-.15	.06	.39
4	-.03	-.08	-.07	-.08	.18	.02	-.15	-.08	-.13	.01
5	.22	.02	-.42	.16	.17	.09	.04	-.14	.02	.31
6	.04	.25	.12	.06	.26	-.08	-.14	-.24	.03	.23
7	.10	.11	.05	.18	-.03	-.23	-.31	.26	.13	.30
8	.40	.43	.15	-.05	-.06	.17	.01	-.16	.02	.43
9	.15	-.31	.05	.03	.34	.04	.00	-.17	.12	.29
10	.32	.40	.01	-.20	-.07	-.24	.18	.04	.08	.41
11	.13	-.36	.40	.01	.00	.00	-.26	.09	.03	.39
12	.35	-.13	.09	-.02	.07	-.08	-.02	-.21	.09	.21
13	.06	-.11	-.45	-.09	-.15	-.15	-.22	-.00	.02	.32
14	.19	.02	-.08	-.39	.05	.04	-.24	-.03	-.20	.30
15	.31	-.09	.05	-.06	-.27	-.31	.03	-.14	-.13	.31
16	.05	.04	.16	.32	-.20	.28	-.08	-.09	-.13	.28
17	.18	-.42	-.12	.12	-.22	-.01	.26	-.06	.07	.36
18	.12	.05	-.28	-.27	.15	.26	.06	.24	.05	.33
19	.05	.31	.11	.28	.02	.06	.06	.12	-.05	.27
20	-.43	.00	-.00	.05	.19	-.26	.20	-.05	-.16	.36
Factor Contribution	1.22	1.02	.92	.76	.66	.54	.53	.44	.21	6.29
% of Common Variance	26.30	22.06	19.79	16.34	14.24	11.67	11.39	9.60	4.50	135.90
% of Total Variance	6.08	5.10	4.58	3.78	3.29	2.70	2.63	2.22	1.04	31.44

TABLE C-14

Principal Factor Loadings Matrix for Test 8 (Random data)

Item	Factor									
	1	2	3	4	5	6	7	8	9	h^2
1	-.56	.12	.21	.02	-.14	.07	.14	-.11	-.04	.43
2	-.29	.02	-.17	-.10	.28	-.28	.12	.13	-.11	.32
3	.10	.32	.08	-.15	-.08	-.03	-.13	-.29	.06	.25
4	.26	.10	-.30	-.28	-.20	.28	.20	.04	.02	.40
5	.04	-.32	-.02	.07	.12	-.01	-.27	-.14	-.26	.28
6	-.05	-.05	-.23	.31	-.34	.04	-.08	.07	.14	.31
7	-.02	-.03	-.26	-.05	-.28	-.23	-.21	-.08	.03	.25
8	-.02	.21	.40	-.18	-.14	.12	.10	.08	-.05	.29
9	.50	.20	-.14	-.06	.26	.26	-.00	.04	-.12	.47
10	.40	.13	-.08	.04	.06	-.11	-.00	-.17	.20	.27
11	-.34	-.02	-.01	-.18	-.09	.26	-.17	-.01	.05	.26
12	-.05	.24	-.07	-.19	-.12	-.12	.14	-.18	-.15	.20
13	-.12	-.12	-.36	-.33	-.10	-.08	-.03	.22	-.02	.34
14	-.11	-.14	-.29	.32	-.15	.25	.22	-.13	-.13	.40
15	.09	.15	.06	-.03	-.33	-.11	-.14	.02	-.25	.27
16	.11	.50	.22	-.29	-.07	.12	-.08	.03	.12	.44
17	-.06	.57	.01	.16	-.03	-.01	-.17	.26	.03	.46
18	-.28	.14	-.24	-.28	.11	.07	-.12	-.13	.03	.28
19	-.02	.02	.08	.07	-.02	.28	-.24	.06	-.11	.17
20	-.32	.18	-.15	.06	.29	.19	-.13	-.08	.08	.31
Factor Contribution	1.27	1.08	.83	.74	.70	.61	.46	.38	.30	6.38
% of Common Variance	25.14	21.48	16.50	14.64	13.96	12.14	9.21	7.50	5.94	126.51
% of Total Variance	6.34	5.42	4.16	3.69	3.52	3.06	2.32	1.89	1.50	31.92

TABLE C-15

Principal Factor Loadings Matrix for Test 9 (Random Data)

Item	Factor									
	1	2	3	4	5	6	7	8	9	h^2
1	-.05	.40	.32	-.19	.12	.12	.03	-.13	-.06	.35
2	.30	.32	-.09	.05	.15	-.04	-.21	.21	-.05	.32
3	-.17	.03	.10	.34	-.07	.20	.12	.06	-.23	.27
4	.16	.14	.18	-.06	-.24	-.32	-.11	-.20	.10	.31
5	.00	.43	.14	.13	.01	-.24	.06	.10	.17	.32
6	-.00	-.15	.29	.28	.27	-.08	.17	-.10	.02	.30
7	-.29	-.18	.12	.14	-.17	.08	-.16	-.10	.22	.27
8	.55	-.07	-.18	-.15	.00	.19	-.10	.03	.09	.42
9	.11	.01	-.35	.31	.19	.13	-.05	-.08	-.02	.30
10	.32	-.22	.12	.13	-.07	-.03	.12	.19	.16	.26
11	-.23	-.13	.21	-.17	.26	.05	-.28	.02	-.05	.29
12	.07	.13	-.08	-.12	-.25	.12	.20	.16	.09	.19
13	-.32	-.46	-.00	-.03	-.03	-.06	-.11	.18	.12	.37
14	.17	-.15	.20	.23	-.03	-.25	.02	.17	-.18	.26
15	-.12	-.13	-.06	-.40	.14	-.21	.06	.15	-.16	.31
16	-.34	.11	-.13	-.10	.19	.09	.25	-.00	.19	.30
17	-.42	.27	-.01	-.00	-.28	.06	-.02	.20	-.08	.37
18	.17	.04	.49	- 04	.05	.32	-.09	.16	.09	.42
19	.14	-.12	.09	-.17	.21	-.05	.25	.05	.06	.18
20	-.22	.24	-.16	.17	.27	-.13	-.12	.15	.18	.32
Factor Contribution	1.24	1.01	.82	.73	.63	.54	.44	.38	.34	6.14
% of Common Variance	26.35	21.56	17.59	15.64	13.38	11.60	9.37	8.05	7.37	130.92
% of Total Variance	6.17	5.05	4.12	3.67	3.14	2.72	2.19	1.89	1.73	30.68

TABLE C-16

Principal Factor Loadings Matrix for Test 10 (Random data)

Item	Factor									
	1	2	3	4	5	6	7	8	9	h^2
1	-.16	-.35	.02	-.03	-.22	-.13	-.21	-.05	.11	.27
2	.14	-.08	-.24	-.18	-.25	-.14	-.06	.35	-.06	.33
3	-.22	.09	.37	.05	.09	-.11	-.02	.19	.03	.26
4	-.10	-.16	.16	.32	-.27	.13	-.13	.03	.10	.28
5	.17	-.05	-.15	.37	.14	-.18	-.20	.07	-.10	.29
6	.06	.22	.06	.22	-.18	-.30	-.07	-.10	.07	.25
7	.07	.23	.10	-.31	.14	-.10	-.19	.07	.19	.27
8	.16	-.27	-.18	.09	.20	.30	-.02	.01	.03	.27
9	.35	.30	-.22	-.03	-.16	-.16	.02	-.18	-.12	.36
10	-.25	.21	-.33	.24	-.04	-.12	.19	.14	.12	.36
11	.46	.14	-.22	.02	-.19	.22	-.01	-.05	.14	.39
12	-.10	.18	.18	-.10	-.09	.16	.03	-.09	-.04	.13
13	.23	-.10	.15	-.22	-.03	-.17	.27	.08	-.06	.25
14	.23	-.16	.03	.19	.23	-.16	.19	.13	.02	.24
15	.21	-.17	.33	.15	-.21	-.02	.09	-.06	-.23	.31
16	-.24	-.14	-.07	-.18	-.08	-.21	-.10	-.12	-.08	.20
17	-.28	-.25	-.14	-.00	-.05	-.14	.28	-.20	.20	.34
18	-.41	.40	-.02	.14	.08	.07	.01	-.02	-.14	.37
19	-.38	-.03	-.21	-.08	-.27	.17	.06	.16	-.14	.35
20	.14	.19	.28	.07	-.25	.10	.12	.11	.17	.27
Factor Contribution	1.19	.87	.80	.66	.62	.57	.41	.37	.30	5.79
% of Common Variance	27.61	20.14	18.54	15.35	14.25	13.10	9.44	8.54	7.02	133.99
% of Total Variance	5.97	4.35	4.01	3.32	3.08	2.83	2.04	1.85	1.52	28.96

TABLE C-17

Principal Factor Loadings Matrix for Test 11 (Random data)

Item	Factor									
	1	2	3	4	5	6	7	8	9	h^2
1	-.31	.25	.30	-.05	-.27	.04	.09	.07	.14	.36
2	.02	.16	-.25	-.36	.21	.06	-.24	-.22	-.08	.38
3	-.24	-.02	.17	.07	.03	-.06	-.45	.16	.11	.33
4	-.35	.30	.22	.09	.02	.16	.08	-.21	.10	.35
5	.30	-.24	.06	.04	.22	.14	.14	.04	.16	.27
6	-.08	-.13	-.17	-.18	-.14	-.32	.15	.03	-.04	.23
7	-.07	-.00	.35	.17	-.06	.07	-.10	-.11	-.13	.20
8	.34	.30	-.01	-.01	-.27	-.08	-.26	.05	.03	.36
9	.18	.09	.09	.33	-.09	-.35	-.07	-.06	-.10	.30
10	-.42	-.00	-.11	-.03	-.29	.03	-.00	-.03	-.13	.29
11	-.21	.42	-.27	.22	.10	-.02	.10	.17	.15	.42
12	.06	.13	-.09	-.31	-.28	.06	.14	-.15	.15	.27
13	.23	-.06	-.28	.34	-.33	.12	.11	.04	-.09	.40
14	.22	.19	-.02	-.34	-.08	.07	-.06	.26	.01	.28
15	-.36	-.20	-.33	.21	.08	.08	-.09	-.05	.10	.35
16	-.00	.07	.13	-.06	-.02	-.22	.05	-.22	-.03	.12
17	.38	.18	.09	.12	.10	-.23	.04	-.13	.22	.33
18	-.26	.35	-.07	-.05	.29	-.28	.16	.13	-.16	.42
19	-.17	-.20	-.33	-.01	-.11	-.25	-.14	-.10	.20	.32
20	.19	.44	-.29	.18	.04	.22	.07	-.12	-.08	.42
Factor Contribution	1.25	1.02	.90	.78	.67	.61	.51	.37	.30	6.42
% of Common Variance	25.90	21.10	18.51	16.11	13.84	12.53	10.47	7.70	6.30	132.48
% of Total Variance	6.27	5.11	4.48	3.90	3.35	3.03	2.54	1.86	1.53	32.08

TABLE C-18

Principal Factor Loadings Matrix for Test 12 (Random data)

Item	Factor									
	1	2	3	4	5	6	7	8	9	h^2
1	.12	-.14	-.19	.00	-.09	.40	-.21	-.02	.09	.29
2	.46	-.02	.02	.23	-.04	-.09	.19	.09	.18	.34
3	.08	-.07	.29	.10	-.31	.14	-.09	-.17	-.13	.28
4	-.01	-.54	-.07	-.20	.15	.03	.12	-.02	.10	.39
5	.42	.04	-.18	-.22	-.27	-.04	.07	.06	-.16	.36
6	-.04	.36	.25	-.11	.02	.02	.21	-.09	.17	.29
7	-.05	-.09	.39	-.07	-.17	-.27	-.06	.04	.09	.28
8	-.11	-.38	.13	-.09	-.10	.20	.10	-.25	.08	.31
9	.44	.01	-.13	-.05	.21	-.08	-.14	-.11	.14	.31
10	.48	.07	-.10	.05	-.04	-.03	-.25	-.04	-.05	.32
11	-.05	-.26	-.02	-.33	-.16	-.15	-.18	.08	-.02	.27
12	.18	-.17	.49	-.06	-.03	.10	-.16	.17	.07	.38
13	.17	.08	.13	-.18	.24	.09	-.18	-.02	.10	.20
14	.02	.07	.31	-.21	.20	-.22	-.08	-.18	-.12	.29
15	-.04	.43	.05	-.06	-.18	.02	-.12	.04	.17	.27
16	-.47	.06	-.06	.15	.01	-.01	-.36	.05	.07	.39
17	-.14	-.07	-.41	-.22	-.04	-.20	-.06	.00	.10	.30
18	-.03	.39	-.13	-.28	-.10	.11	.02	-.21	.01	.32
19	.04	.11	.10	-.10	.33	.21	-.02	.12	-.16	.23
20	.08	-.13	-.04	.30	.07	-.19	-.15	-.27	-.01	.25
Factor Contribution	1.16	1.08	.98	.62	.57	.53	.52	.33	.26	6.05
% of Common Variance	24.97	23.42	21.14	13.37	12.32	11.47	11.25	7.16	5.58	130.69
% of Total Variance	5.78	5.42	4.89	3.09	2.85	2.66	2.61	1.66	1.29	30.25

DISTRIBUTION LIST

Navy

4 Dr. Marshall J. Farr, Director
Personnel & Training Research Programs
Office of Naval Research
Arlington, VA 22217

1 Director
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Director
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101
ATTN: E. E. Gloye

1 Director
ONR Branch Office
536 South Clark Street
Chicago, IL 60605
ATTN: M. A. Bertin

6 Director
Naval Research Laboratory
Code 2627
Washington, DC 20390

12 Defense Documentation Center
Cameron Station, Building 5
5010 Duke Street
Alexandria, VA 22314

1 Chairman
Behavioral Science Department
Naval Command and Management Division
U.S. Naval Academy
Luce Hall
Annapolis, MD 21402

1 Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054
ATTN: Dr. N.J. Kerr

1 Chief of Naval Training
Naval Air Station
Pensacola, FL 32508
ATTN: CAPT Bruce Stone, USN

1 LCDR Charles J. Theisen, Jr., MSC, USN
4024
Naval Air Development Center
Warminster, PA 18974

1 Commander
Naval Air Reserve
Naval Air Station
Glenview, IL 60026

1 Mr. Lee Miller (AIR 413E)
Naval Air Systems Command
5600 Columbia Pike
Falls Church, VA 22042

1 Dr. Harold Booher
NAVAIR 415C
Naval Air Systems Command
5600 Columbia Pike
Falls Church, VA 22042

1 CAPT John F. Riley, USN
Commanding Officer
U.S. Naval Amphibious School
Coronado, CA 92155

1 Special Assistant for Manpower
OASN (M&RA)
The Pentagon, Room 4E794
Washington, DC 20350

1 Dr. Richard J. Niehaus
Office of Civilian Manpower Management
Code 06A
Department of the Navy
Washington, DC 20390

1 CDR Richard L. Martin, USN
COMFAIRMIRAMAR F-14
NAS Miramar, CA 92145

1 Research Director, Code 06
Research and Evaluation Department
U.S. Naval Examining Center
Great Lakes, IL 60088
ATTN: C. S. Winiewicz

1 Chief
Bureau of Medicine and Surgery
Code 413
Washington, DC 20372

1 Commanding Officer
Naval Medical Neuropsychiatric
Research Unit
San Diego, CA 92152

1 Technical Reference Library
Naval Medical Research Institute
National Naval Medical Center
Bethesda, MD 20014

1 Chief
Bureau of Medicine and Surgery
Research Division (Code 713)
Department of the Navy
Washington, DC 20372

1 Dr. John J. Collins
Chief of Naval Operations (OP-987F)
Department of the Navy
Washington, DC 20350

1 Technical Library
Bureau of Naval Personnel
Department of the Navy
Washington, DC 20360

1 Head, Personnel Measurement Staff
Capital Area Personnel Office
Ballston Tower #2, Room 1204
801 N. Randolph Street
Arlington, VA 22203

1C Dr. James J. Regan, Technical Director
Navy Personnel Research
and Development Center
San Diego, CA 92152

1 Dr. Norman Abrahams
Navy Personnel Research
and Development Center
San Diego, CA 92152

1 Dr. Bernard Rimland
Navy Personnel Research
and Development Center
San Diego, CA 92152

1 Superintendent
Naval Postgraduate School
Monterey, CA 92940
ATTN: Library (Code 2124)

1 Technical Library
Naval Ship Systems Command
National Center, Building 3
Room 3S08
Washington, DC 20360

1 Commanding Officer
Service School Command
U.S. Naval Training Center
San Diego, CA 92133
ATTN: Code 303

1 Chief of Naval Training Support
Code N-21
Building 45
Naval Air Station
Pensacola, FL 32508

1 Dr. William L. Maloy
Principal Civilian Advisor
for Education and Training
Naval Training Command, Code 01A
Pensacola, FL 32508

1 Dr. Hanss H. Wolff
Technical Director (Code N-2)
Naval Training Equipment Center
Orlando, FL 32813

1 CDR Fred Richardson
Navy Recruiting Command
BCT #3, Room 215
Washington, DC 20370

1 Mr. Arnold Rubinstein
Naval Material Command (NMAT-03424)
Room 820, Crystal Plaza #6
Washington, DC 20360

1 Dr. H. Wallace Sinaiko
c/o Office of Naval Research (Code 450)
Psychological Sciences Division
Arlington, VA 22217

1 Director, Navy Occupational Task
Analysis Program (NOTAP)
Navy Personnel Program Support
Activity
Building 1304, Bolling AFB
Washington, DC 20336

Army

1 Commandant
U.S. Army Institute of Administration
ATTN: EA
Fort Benjamin Harrison, IN 46216

1 Armed Forces Staff College
Norfolk, VA 23511
ATTN: Library

1 U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

1 Commanding Officer
ATTN: LTC Montgomery
USACDC - PASA
Ft. Benjamin Harrison, IN 46249

1 U.S. Army Research Institute
Commonwealth Building, Room 239
1300 Wilson Boulevard
Arlington, VA 22209
ATTN: Dr. R. Dusek

1 Mr. Edmund F. Fuchs
U.S. Army Research Institute
1300 Wilson Boulevard
Arlington, VA 22209

1 Chief, Unit Training and Educational Technology Systems
U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

1 Dr. Stanley L. Cohen
Work Unit Area Leader
Organizational Development Work Unit
Army Research Institute for Behavioral and Social Science
1300 Wilson Boulevard
Arlington, VA 22209

1 Dr. Leon H. Nawrocki
U.S. Army Research Institute
Rosslyn Commonwealth Building
1300 Wilson Boulevard
Arlington, VA 22209

Air Force

1 Headquarters, U.S. Air Force
Chief, Personnel Research and Analysis Division (AF/DPSY)
Washington, DC 20330

1 Research and Analysis Division
AF/DPXYR Room 4C200
Washington, DC 20330

1 AFHRL/AS (Dr. G. A. Eckstrand
Wright-Patterson AFB
Ohio 45433

1 AFHRL/MD
701 Prince Street
Room 200
Alexandria, VA 22314

1 Dr. Robert A. Bottenberg
AFHRL/PES
Lackland AFB, TX 78236

1 Personnel Research Division
AFHRL
Lackland Air Force Base
Texas 78236

1 AFOSR(NL)
1400 Wilson Boulevard
Arlington, VA 22209

1 Commandant
USAF School of Aerospace Medicine
Aeromedical Library (SUL-4)
Brooks AFB, TX 78235

1 CAPT Jack Thorpe, USAF
Department of Psychology
Bowling Green State University
Bowling Green, OH 43403

1 Lt. Col. Henry L. Taylor, USAF
Military Assistant for Human Resources
OAD(E&LS) ODDR&E
Pentagon, Room 3D129
Washington, DC 20301

Marine Corps

1 Commandant, Marine Corps
Code AO1M-2
Washington, DC 20380

1 COL George Caridakis
Director, Office of Manpower Utilization
Headquarters, Marine Corps (AO1H)
MCB
Quantico, VA 22134

1 Dr. A. L. Slafkosky
Scientific Advisor (Code Ax)
Commandant of the Marine Corps
Washington, DC 20380

1 Mr. E. A. Dover
Manpower Measurement Unit (Code AO1M-2)
Arlington Annex, Room 2413
Arlington, VA 20370

Coast Guard

1 Mr. Joseph J. Cowan, Chief
Psychological Research Branch (P-1)
U.S. Coast Guard Headquarters
400 Seventh Street, SW
Washington, DC 20590

Other DOD

1 Lt. Col. Austin W. Kibler, Director
Human Resources Research Office
Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Other Government

1 Dr. Lorraine D. Eyde
Personnel Research and Development Center
U.S. Civil Service Commission, Room 3458
1900 E. Street, N.W.
Washington, DC 20415

1 Dr. Vern Urry
Personnel Research and Development
Center
U.S. Civil Service Commission
Washington, DC 20415

1 Office of Computer Information
Institute for Computer Sciences
and Technology
National Bureau of Standards
Washington, DC 20234

1 Dr. Eric McWilliams, Program Manager
Technology and Systems, TIE
National Science Foundation
Washington, DC 20550

Miscellaneous

1 Dr. Scarvia B. Anderson
Educational Testing Service
17 Executive Park Drive, N.E.
Atlanta, GA 30329

1 Dr. Richard C. Atkinson
Stanford University
Department of Psychology
Stanford, CA 94305

1 Dr. Bernard M. Bass
University of Rochester
Management Research Center
Rochester, NY 14627

1 Dr. Ronald P. Carver
American Institutes for Research
8555 Sixteenth Street
Silver Spring, MD 20910

1 Century Research Corporation
4113 Lee Highway
Arlington, VA 22207

1 Dr. Kenneth E. Clark
University of Rochester
College of Arts and Sciences
River Campus Station
Rochester, NY 14627

1 Dr. Allan M. Collins
Bolt Beranek and Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Rene' V. Dawis
University of Minnesota
Department of Psychology
Minneapolis, MN 55455

1 Dr. Norman R. Dixon
Associate Professor of Higher Education
University of Pittsburgh
617 Cathedral of Learning
Pittsburgh, PA 15213

1 Dr. Robert Dubin
University of California
Graduate School of Administration
Irvine, CA 92664

1 Dr. Marvin D. Dunnette
University of Minnesota
Department of Psychology
N492 Elliott Hall
Minneapolis, MN 55455

1 Dr. Victor Fields
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Edwin A. Fleishman
American Institutes for Research
8555 Sixteenth Street
Silver Spring, MD 20910

1 Dr. Duncan N. Hansen
Memphis State University
Bureau of Educational Research and Services
Memphis, TN 38152

1 Dr. Robert Glaser, Director
University of Pittsburgh
Learning Research and Development Center
Pittsburgh, PA 15213

1 Dr. Albert S. Glickman
American Institutes for Research
8555 Sixteenth Street
Silver Spring, MD 20910

1 Dr. Harry H. Harman
Educational Testing Service
Division of Analytical Studies and Services
Princeton, NJ 08540

1 Dr. Richard S. Hatch
Decision Systems Associates, Inc.
11428 Rockville Pike
Rockville, MD 20852

1 Dr. M. D. Havron
Human Sciences Research, Inc.
Westgate Industrial Park
7710 Old Springhouse Road
McLean, VA 22101

1 Human Resources Research Organization
Division #3
P.O. Box 5787
Presidio of Monterey, CA 93940

1 Human Resources Research Organization
Division #6, Library
P.O. Box 428
Fort Rucker, AL 36360

1 Dr. Lawrence B. Johnson
Lawrence Johnson and Associates, Inc.
200 S Street, N.W., Suite 502
Washington, DC 20009

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. E. J. McCormick
Purdue University
Department of Psychological Sciences
Lafayette, IN 47907

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Drive
Santa Barbara Research Park
Goleta, CA 93017

1 Mr. Edmond Marks
109 Grange Building
Pennsylvania State University
University Park, PA 16802

1 Dr. Andrew R. Molnar
Technological Innovations in Education
National Science Foundation
Washington, DC 20550

1 Dr. Leo Munday
Vice President
American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Mr. Luigi Petrullo
2431 North Edgewood Street
Arlington, VA 22207

1 Dr. Diane M. Ramsey-Klee
R-K Research & System Design
3947 Ridgemont Drive
Malibu, CA 90265

1 Dr. Joseph W. Rigney
Behavioral Technology Laboratories
University of Southern California
3717 South Grand
Los Angeles, CA 90007

1 Dr. George E. Rowland
Rowland and Company, Inc.
P.O. Box 61
Haddonfield, NJ 08033

1 Dr. Arthur I. Siegel
Applied Psychological Services
Science Center
404 East Lancaster Avenue
Wayne, PA 19087

1 Mr. Dennis J. Sullivan
725 Benson Way
Thousand Oaks, CA 91360

1 Dr. Benton J. Underwood
Northwestern University
Department of Psychology
Evanston, IL 60201

1 Dr. Anita West
Denver Research Institute
University of Denver
Denver, CO 80210

1 Dr. John Annett
The Open University
Milton Keynes
Buckinghamshire
ENGLAND

1 Dr. Milton S. Katz
MITRE Corporation
Westgate Research Center
McLean, VA 22101

1 Dr. Charles A. Ullmann
Director, Behavioral Sciences Studies
Information Concepts, Inc.
1701 N. Ft. Myer Drive
Arlington, VA 22209

1 Dr. Alfred F. Smode, Staff Consultant
Training Analysis and Evaluation Group
Naval Training Equipment Center
Code N-00T
Orlando, FL 32813